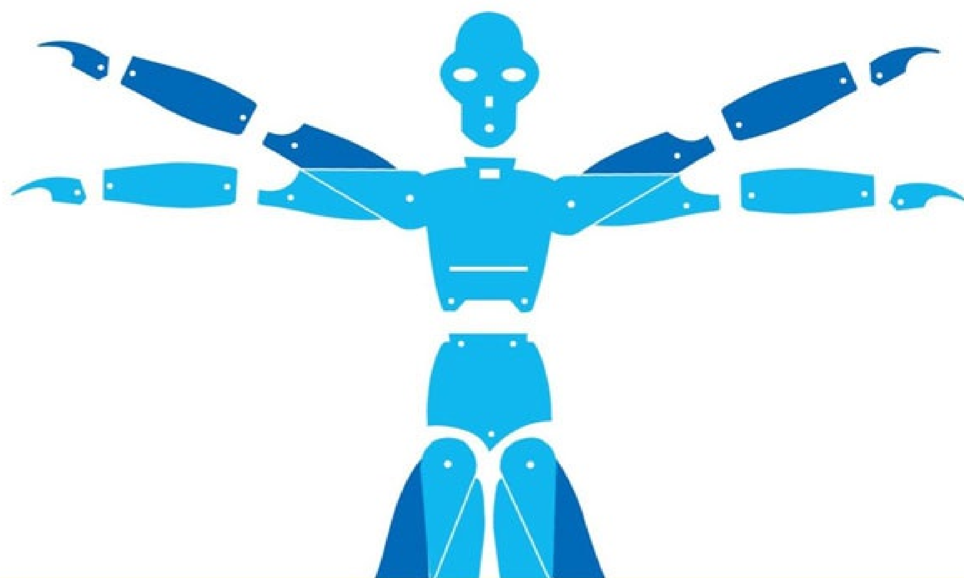
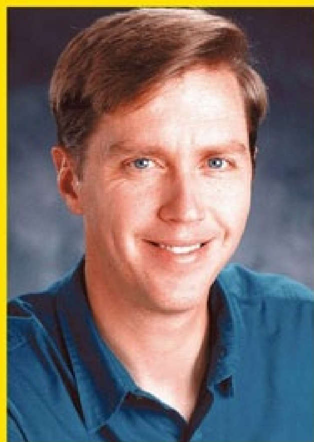
ON INTELLIGENCE

智能时代

当所有的机器都能学习思考，我们的生活会如何改变

[美] 杰夫·霍金斯
桑德拉·布拉克斯莉 ◎著
李蓝 刘知远◎译





世界科技产业界领袖级人物、掌上电脑PDA发明人

杰夫·霍金斯 经典力著全新升级版

详细揭示未来主流大趋势，比大数据更能决定我们生活的是智能
可穿戴设备、智能手机、智能汽车、智能家居……智能时代已经大踏步来临

深入人脑核心区域，探究人类智能原理
两届诺贝尔奖得主强烈推荐

中国華僑出版社

版权信息

书名:智能时代

作者:[美]杰夫·霍金斯 桑德拉·布拉克斯莉

译者:李蓝 刘知远

中信出版集团制作发行

引言 制造出像大脑一样工作的机器

这本书的创作，连同我的生活，由两种激情共同驱动着。

近25年来，我一直热衷并投身于移动计算机领域。在硅谷的高科技世界里，我因创办了Palm Computing和Handspring两家公司而声名在外。作为掌上电脑和智能手机的架构设计师，我还曾设计出PalmPilot和Treo一类的产品。

然而，我的第二种兴趣不但早于我对计算机的热情，对我个人来说也更为重要——我疯狂地着迷于大脑。我渴望了解大脑是如何工作的——不单是从哲学的角度和笼统的概念上去理解，还要采用工程学的方式，从最细节处去彻底掌握。光是了解“智能是什么”以及“大脑是如何工作的”并不足够，我还想要知道，如何才能制造出像大脑那样工作的机器。总而言之，我想要创造真正具有智能的机器。

有关智能的问题，构成了科学界最后一片壮阔的前沿领域。大多数重大的科学问题所涉及的事件，往往极为微小或极为庞大，有的甚至发生在遥远的亿万年前。然而，有关智能的问题却与人类切身相关。人人都有一颗大脑，你的大脑即是你本人。你为何会产生这样而不是那样的感觉？你如何感知世界？为何你会犯错？如何才能变得富有创意？为什么音乐和艺术能够激发灵感？生而为人，究竟意味着什么？若想找寻这些问题的答案，我们首先必须了解大脑。此外，一个能够解释智能和大脑功能的成功理论，不单能够帮助我们治疗与大脑有关的疾患，还将带来巨大的社会效益。据此理论建造出的真正智能机器，绝不同于通俗小说和科学幻想中所描绘的那种。相反，这些智能机器将从一套关于智能本质的新理论中诞生。它们将帮助人类加速认识世界、探索宇宙，令世

界更加和平。一个相关的大型产业，也将在此过程中逐渐形成。

幸运的是，我们生于一个有望解答智能问题的时代。我们这一代人，拥有着数百年来收集的堆积如山的关于大脑的数据，而数据收集的速度还在日益加快。仅在美国就有成千上万的神经科学家。然而，学术界至今也没有形成一套能够富有成效地解释智能本质或大脑工作原理的完整理论。大部分的神经生物学家并不关心有关大脑的整体理论，他们沉醉于做实验来收集更多有关大脑诸多子系统的数据。尽管计算机程序员们前仆后继地尝试让计算机拥有智能，但这些努力最后均以失败告终。我相信，如果他们继续对电脑和人脑之间的差别视而不见，失败也还将继续下去。

智能究竟是什么？为何它只存在于人脑，而不存在于电脑？为什么一个6岁的孩子能够姿态优美地在河床中的岩石上来回跳跃，而我们这个时代最先进的机器人走起路来却像是行动笨拙的僵尸？为什么人类3岁的小孩就已经能够以自己的方式掌握语言，而程序员们耗费了近半个世纪的心血，仍不能让计算机实现同样的成就？为什么你能在不到一秒的时间内准确分辨出猫和狗，而超级计算机却做不到？这些都是等待我们去探索的伟大奥秘。我们已经拥有了大量的线索，而现在真正需要的，是一些关键的启迪。

你可能会奇怪，为什么一个计算机设计师在写一本关于大脑的书。换句话说，既然我真的那么热爱大脑，为什么不选择脑科学或者人工智能研究作为自己的职业呢？答案是，我试过了，而且不止一次。但是我无法接受像前人那样研究智能问题。我相信解决这一问题的最好办法，是以大脑的生物学细节作为约束和指导，同时将智能视为一个计算性质的问题——将其定位于生物学和计算机科学之间。许多生物学家拒绝或忽视在计算机的语境下去思考大脑，而计算机科学家们通常也不相信能从生物学中得到任何可借鉴之处。而且，科学界比商界更不愿承担风险。在科技行业，如果一个人以合理的方法追求新想法的实现，无论最

后成功与否，都将促进自己事业的发展。许多成功的企业家早期都品尝过失败的滋味。但在学术界，如果对一个新想法投入的心血在几年之后仍不见成果，你的职业生涯很可能就此夭折。正因如此，我决定同时追求生命中的这两种激情，并坚信商业上的成功将有利于我取得有关大脑研究上的成功。我需要财力来支持我的科学追求，同时我也需要学习影响世界和推销新想法的方法，所有这一切，当时我都希望能从在硅谷的工作中获得。

2002年8月，我创办了一家名为“红杉神经科学研究所”（Redwood Neuroscience Institute, RNI）的研究中心，从事大脑理论的研究。世界上的神经科学中心有许多个，但专门致力于为新皮层寻找全面解释理论的，仅此一家。而新皮层正是大脑中负责智能的部分。这就是我们在RNI的全部研究课题。从许多方面来看，RNI就像是一家初创公司。我们正追逐着在某些人看来遥不可及的梦想，幸运的是，我们的团队卧虎藏龙，大家的努力已初见成果。

* * *

这本书的议题称得上是雄心勃勃。它要提出一个描绘大脑如何工作的全面理论，包括什么是智能，以及大脑如何创造智能。我所提出的这个理论并不是全新的。你将要读到的许多想法，都曾以这样或那样的形式散落各处，却从未被以连贯的方式串联起来，而这正是本书的首创。这也不足为奇，“新想法”往往是旧想法的重新包装和重新诠释。这句话的确也适用于本书中所提出的理论。但包装和诠释能令新旧想法产生天壤之别——即一堆繁杂细节和一个能够令人满意的理论之间的差别。这个理论已经打动了很多人，我希望它也能打动你。我听到的一种典型的反应是：“有道理，我本来不会想到从这个角度来看智能，但听你描述完之后，我能理解这是怎么一回事了。”一旦拥有了这方面的知识，大部分人会开始用不同的眼光看待自己。你将开始观察自己的行为：“我明白刚刚在我的大脑里发生了什么。”希望读完这本书的时候，你对于

自己所思所行的缘由，能够有一个新的认识。我也希望某些读者能自书中得到启发，并且根据书中的原理，致力于建造智能机器的事业。

通常，我会把这个理论和我研究智能的方式称为“真正的智能”（real intelligence），以区别于“人工智能”。人工智能科学家试图通过编程，让计算机表现得像人类，却没有先回答智能是什么，其含义又是什么。他们遗漏了建造智能机器最重要的部分——智能！而“真正的智能”则认为，在尝试构建智能机器之前，我们必须首先了解大脑是如何思考的，这里并没有丝毫人工的东西。只有到那时，我们才可以问，如何能够建造智能机器。

本书在前五章首先介绍为何先前在理解和建造智能机器方面的努力均告失败，随后提出并进一步论述我称之为“记忆－预测架构”（memoryprediction framework）理论的核心概念。第六章则详述了物质大脑如何实现记忆－预测模型，换句话说，就是大脑实际上是如何工作的。接下来的第七章讨论了这个理论对社会和其他方面的影响，对许多读者来说，这可能是本书中最引人深思的部分。最后，本书以围绕智能机器的讨论作为结束，重点探讨人类如何能建造智能机器，以及它的未来将是什么样子。希望你会为之神往。以下是我们即将一一探讨的问题：

计算机能够拥有智能吗？

数十年来，人工智能领域的科学家们宣称，当计算机足够强大时，就将拥有智能。我不这么认为，后面我会解释为什么。大脑和计算机的工作原理根本就是两回事。

神经网络方面的研究能否导致智能机器的产生？

大脑固然是由神经元网络构成的，但如果不理解大脑的工作原理，仅凭简单的神经网络研究，在创造智能机器方面绝不会比计算机编程更有优势。

理解大脑的工作原理为何如此困难？

大多数科学家认为，大脑太过于复杂，因此需要很长的时间才能理解它。我不同意这个观点。复杂是思维混乱的表现，而不是其原因。相反，我认为，我们所持有的一些直觉假设误导了我们。这其中最大的错误，就是将智能等同于表现出智能的行为。

如果智能不由行为定义，那么该如何定义它？

大脑使用大量的记忆资源来创建关于世界的模型。你所知道和所学到的一切，都存储在这个模型中。大脑根据这个基于记忆的模型，不断对未来事件作出预测。预测未来的能力，才是定义智能的关键。我将深入描述大脑的预测能力，它正是本书的核心概念。

大脑是如何工作的？

智能产生于大脑的新皮层。尽管拥有诸多能力和极强的适应性，新皮层的结构细节却出奇地规则。新皮层的不同部位，无论是负责视觉、听觉、触觉，还是语言的部分，都遵循着相同的工作原理。理解新皮层的关键就在于理解这些共同的原理，尤其是它们的层级结构。我们将从详尽的细节入手来考察新皮层，为你展示它如何用自身结构来捕获这个世界的结构。这些讨论将会是本书中最具技术性的部分，但对于对此感

兴趣的非科学家读者们来说，也不难理解。

这个新理论有什么意义？

这个大脑理论可以帮助我们解释许多事情，比如我们怎样才能有创意？我们为什么会拥有意识？我们为何产生偏见？我们是如何学习的？以及为什么说“老狗学不会新把戏”，等等。我将会讨论许多这类话题。总之，这一理论能帮助我们认识自己，并了解我们自身行为产生的原因。

我们有能力建造出智能机器吗？它们能做什么？

是的，我们能够而且我们将会建造出智能机器。可以预见的是，在今后的数十年，这种机器的性能将朝着许多有趣的方向迅速发展。有些人担心智能机器可能会在未来危害人类的生存，对此无稽的想法我表示强烈反对。人类并不会被机器人超越。建造在物理、数学等高层次认知能力上超过我们的智能机器，要比建造科幻小说中的那种会走路、会说话的机器人容易得多。我将探讨建造智能机器的技术可能的一些发展方向，那将是令人难以置信的方向。

我的目标，就是以人人都能理解的方式，来说明这个新的智能理论和大脑的工作原理。一个好的理论应该易于理解，而不是掩藏于一堆艰涩的术语和错综复杂的论述中。我将首先从对大脑基本架构的介绍出发，并随着我们讨论的深入，逐步延伸。有些细节纯粹是逻辑推论，有些则会涉及大脑神经回路的特定方面。某些细节难免会被证明有错，不过这一点在任何科学领域都不可避免。一个完全成熟的理论需要长年累月的发展，但核心概念的力量并不会因此而有丝毫减损。

* * *

许多年前，当我第一次对大脑发生兴趣的时候，我去了当地图书馆，想要找一本解释大脑如何工作的书。当时我只有十几岁，已经习惯于从图书馆中找到写作精良的解释各种有趣话题的好书，比如那时我所着迷的相对论、黑洞、魔术和数学。然而，当我想要寻找一本令人满意的解释大脑的好书时，这一愿望却落了个空。我逐渐意识到，没有人知道大脑实际上是如何工作的，甚至连一个糟糕的或者未经证实的理论都不存在，一片空白。这太出乎意料了。好比说，那时虽然没有人知道恐龙是如何灭绝的，但却有大量的理论供你阅读。而大脑的情况完全不同。起初我很难相信，人们居然不知道这样一个重要器官是如何工作的。这一事实令我倍感困扰。在对大脑的已知知识的学习过程中，我逐渐开始相信，一定存在着一个简单、直观的解释。大脑不是魔术，我认为答案也不会比魔术更复杂。数学家保罗·埃尔德什（Paul Erdos）相信，最简单的数学证明早已存在于宇宙的“天书”中，而数学家的任务就是要去解读“天书”并找出它们。同样，我认为对于智能的解释就在“那儿”。我能感受到它的存在，我要解读这本“天书”。

25年来，写作一本简单、直观地解释大脑的小书，一直是我的愿望。它就像一根悬于眼前的胡萝卜，激励着我不断前进。这一愿望最后变成了你手中的这本书。我一向讨厌复杂，无论是在科学上还是在技术上。这一点从我所设计的产品中就可以看出来，它们往往以简单易用取胜。强大的事物往往是简单的。本书所提出的正是一个简明直观的智能理论，希望你会喜欢。

第一章 人工智能

1979年6月，我从康奈尔大学顺利毕业，并拿到了电机工程专业的学士学位。当时的我对人生还没有任何像样的打算。我在位于俄勒冈州波特兰市的英特尔工业园找了一份工程师的工作。那时，微型计算机产业如日方升，而英特尔公司正处于该领域的核心位置。单板机是英特尔当时的主要产品（得益于英特尔发明的微处理器，把整个计算机置于单个电路板上的想法在当时刚刚成为可能），而我的工作，就是分析并解决由其他工程师们发现的单板机上的问题。期间，我发表了一篇业务通讯。由于常常在外奔波，我有机会结识到各种各样的客户。尽管很思念在辛辛那提工作的大学女友，但那时的我还年轻，日子过得很快活。

几个月后遇到的一件事，改变了我的人生方向。那是当年9月新出版的一期《科学美国人》，一整本都是对大脑研究的介绍。这期引人入胜的杂志，重新点燃了我在少年时代就萌发了对大脑的兴趣。我从中了解到了大脑的组织、发展和化学特征，还有视觉、运动和其他人类专长的神经机制，以及精神失常的生物学基础。这是有史以来最棒的几期《科学美国人》之一。一些与我交流过的神经学家告诉我，这期杂志对他们的职业选择影响巨大——对我也是一样。

该期杂志的最后一篇文章——《关于大脑的思考》，由DNA结构的发现者弗朗西斯·克里克撰写。他那时已经转投大脑研究领域。克里克认为，尽管科学家们对于大脑的细节知识已有大量积累，但大脑的工作原理仍然是一个极大的谜题。科学家通常会避免去写他们所不知道的事物，但克里克就像那个指出皇帝没有穿衣服的小男孩一样，对此毫无顾忌。克里克宣称，神经科学只是一堆没有任何理论的庞大数据。他的原话是：“明显缺乏一个大的理论框架。”在我听来，这就像是一位英国绅

士以委婉的措辞表示：“我们对此事完全摸不着头脑。”在当时，这是句大实话，即便到了今天，也仍然不假。

克里克的话就像是一声集结号，将我研究大脑和建造智能机器的人生梦想就此唤醒。虽然当时我已远离了学校，但还是毅然决定转行。我开始计划研究大脑，不光要了解它的工作原理，还要以这些知识为基础，发展新技术来构建智能机器。但将这一计划付诸行动，仍需要一些时间。

1980年的春天，我被调至英特尔驻波士顿的分部，与我后来的妻子团聚了。她那时刚开始攻读研究生。我一边负责给客户和员工讲解如何设计微处理器系统，一边已将目光投向了另外的目标：寻找发展大脑理论的办法。工程师的直觉告诉我，一旦破解了大脑的工作原理，便可以照葫芦画瓢地造出它们，而用来建造人工大脑的天然材料就是硅。考虑到我所供职的英特尔公司发明了硅内存芯片和微处理器，我想也许能说服公司，允许我将部分时间用于有关智能的研究和设计仿生大脑的记忆芯片上。于是我给英特尔当时的董事长——戈登·摩尔（Gordon Moore）去了一封信，内容大致如下：

亲爱的摩尔博士：

我提议我们公司成立一个专门研究大脑工作原理的小组。该工作可先从一个人（即我本人）开始，以后逐步发展壮大。我有信心我们能干成这事。相信在未来，它将给咱们带来巨大的商机！

杰夫·霍金斯

摩尔看了信后，把我介绍给了英特尔公司的首席科学家特德·霍夫（Ted Hoff）。我飞到加州同他会面，并重申了我的建议——研究大脑。霍夫在两个领域作出了卓越的贡献：一是设计发明了第一台微处理器，二是对早期的神经网络理论作出了出色研究。当时我仅知道前一

桩，而对后一桩并不知晓。霍夫曾在人造神经元及其应用潜力方面进行过研究，但我对这一情况毫无准备。听完我的建议后，他表示并不相信人类在可预见的将来能够弄清大脑的工作原理，因此没有道理让英特尔支持我。我不得不说，他是对的，因为25年后的今天，我们才刚刚开始在理解大脑的道路上有所迈进。在商界，时机决定着一切。不过当时我仍然感到很失落。

我原打算寻求一条捷径来实现目标，在英特尔公司研究大脑本该是最便捷的。既然此路不通，我只好退而求其次，决定去申请麻省理工学院（Massachusetts Institute of Technology, MIT）的研究生。该学院一向以在人工智能方面的研究著称，这对我将来的研究之路大有助益。初看之下，我似乎完全具备申请条件：在计算机科学方面接受过广泛训练——符合；渴望建造智能机器——完全符合；想先从研究大脑开始，了解它的工作原理……呃，等一下，好像有哪里不对了。这最后一项——弄清大脑如何工作，在MIT人工智能实验室的科学家眼里，是“不可能完成之任务”。

我就仿佛一头撞在了南墙上。当时的MIT是人工智能领域的航母。在我申请的同时，大批有才华的头脑正向这里汇聚，着迷于用计算机编程来产生智能行为。对于这些科学家来说，视觉、语言、机器人学和数学都只是编程的问题。既然计算机能够做到大脑所做的一切，甚至更胜一筹，那么我们何必再用大脑这纷繁芜杂的生物性机理来束缚住我们的思想呢？他们认为研究大脑会限制思想，而更好的办法是研究数字计算机所能表现的计算能力的极限。他们的“圣杯”便是编写能够与人脑媲美并最终超越人脑的计算机程序。他们采取了一种以结果来验证手段的研究方法，而对真实的大脑如何工作毫不关心，有些人甚至为自己绕开了神经生物学而沾沾自喜。

正是这一解决问题的错误方式深深震动了我。直觉上，我认为人工智能的方法不仅无法创造出能模仿人类行为的程序，它甚至也不能告诉

我们智能是什么。计算机和人脑的工作原理完全不同，一个只是被编出来的程序，另一个则拥有自我学习的能力；一个必须做到绝对完美才能运行，另一个则天生能够灵活应对，对失误有容忍度；一个具有中央处理器，另一个则不存在中央控制。类似的差别不胜枚举。我之所以认为计算机无法实现智能，最重要的原因是由于我从晶体管层面上理解计算机的工作原理，而这些知识令我产生了一种直觉，即大脑与计算机在本质上完全不同。我无法证明这一直觉，但它之于我，就像人们凭直觉所能确定的任何事一样。最终我得出了结论：人工智能领域的研究或许会催生出有用的产品，但它绝不会建造出真正意义上的智能机器。

与人工智能相反，我想要做的是了解真正的智能和知觉，研究大脑的生理学和解剖学基础，响应弗朗西斯·克里克所提出的挑战，构建一个能解释大脑工作原理的大框架。我把目光投向了大脑的新皮层——哺乳动物的大脑在进化中最新发展出的部分，也是智能的产生之处。只有在认识了新皮层的工作原理之后，我们才能够开始建造智能机器。

遗憾的是，我在MIT的教授和同学中从未遇到知音。他们认为，理解智能和建造智能机器，根本不需要去研究真正的大脑——他们确实是这么说的。1981年，MIT拒绝了我的申请。

* * *

今天，许多人认为，人工智能领域仍然是一片生机，只待计算机拥有足够的计算能力，便可以兑现其许下的种种承诺。按照这一思路，一旦计算机拥有足够的内存和强大的处理器，人工智能的程序员们便能够制造出智能机器。我不同意这一观点。人工智能领域存在着一个根本性的错误：它无法充分解释智能是什么，也无法回答理解事物的能力究竟意味着什么。简要回顾一下人工智能的历史及其建立时的信条，我们就能够看出它是如何偏离正确轨道的。

人工智能的方法是伴随数字计算机的出现而诞生的。早期人工智能

运动中的关键人物，英国数学家阿兰·图灵（Alan Turing），也是“通用计算机”这一想法的提出者之一。他的卓越贡献在于正式提出了“通用计算”的概念：尽管建构的细节存在差异，但从原理上来说，所有的计算机都是相同的。他设想了一个虚拟机器来证明这一点。这个机器由三部分组成：一个处理盒、一条纸带以及一种能从纸带上读取和写入标记、并能来回移动的装置。纸带是用来存储信息的——就像大家熟知的计算机代码1和0一样（当时内存和硬盘还未发明，图灵想象用纸带来存储信息）。处理盒，即今天所谓的中央处理器（CPU），遵循一套固定的规则从纸条上读取和编辑信息。图灵从数学层面上证明，如果为CPU选择了一套正确的规则，并给予它无限长的纸带，这台机器便能够运行宇宙中任何可定义的操作。它可能相当于现在被称为“通用图灵机”的众多机器之一。无论是计算平方根、计算弹道轨迹、玩游戏、编辑照片，还是银行对账，其底层皆是1和0代码，任何一台图灵机都可以通过编写程序来解决这些问题。无论采用何种形式，信息处理就是信息处理。所有的数字计算机在符号逻辑上是等价的。

图灵结论的正确性无可辩驳，在应用上亦富有成效。计算机革命及其所有成果都以此为基石。随后，图灵转向了如何建造智能机器这一问题。他一方面感到计算机可以拥有智能，另一方面却不愿被卷入对其可能性的争论之中。由于认为自己无法给智能一个正式的定义，他甚至没有作出尝试。取而代之，他提出了一个证明智能存在的方法，即著名的图灵测试：如果一台计算机能够骗过人类询问者，诱使他相信它也是人类，那么从定义上来说，这台计算机就拥有智能。以此测试作为检测工具，以图灵机作为媒介，图灵就这样帮助开创了人工智能领域。这一领域的信条是：大脑不过是另一种类型的计算机。因此，只要能让人工智能系统产生与人类相似的行为即可，如何设计它并不重要。

人工智能的拥护者们看到了计算机和人类思维之间的相似之处。他们说：“你看，人类智能行为中最令人印象深刻的，莫过于对抽象符号的操纵——这一点计算机也可以。当我们在说和听的时候，大脑里发生

了什么？我们正是在使用恰当的语法规则操纵着被称为词的心理符号。下棋时呢？是在使用抽象的心理符号来表征各个棋子的属性和位置。看东西时又是怎样呢？我们用抽象心理符号来表征对象的位置、名称和其他属性。当然，人们用以完成这一切的是大脑，而不是我们所建造的各种计算机，但图灵已经表明，如何实现或操纵这些符号，并不重要。你可以使用齿轮、电子开关系统或大脑中的任何神经元网络做到这一点，只要你的媒介可以实现与通用图灵机相同的功能。”

这一假设得到了一篇极富影响力的科学论文的支持。该论文于1943年由神经生理学家沃伦·麦卡洛奇（Warren McCulloch）和数学家沃尔特·皮茨（Walter Pitts）共同发表。在论文中，他们描述了神经元如何能够执行数字功能——神经细胞如何以可理解的方式复制计算机的核心形式逻辑。该想法认为，神经元能够像电脑工程师们所谓的逻辑门一样工作。逻辑门能够实现简单的逻辑运算例如and、not和or。计算机芯片即是由数百万逻辑门连接在一起而组成的精确、复杂的电路。CPU就是逻辑门的集合。

麦卡洛奇和皮茨指出，神经元也可以被连接在一起，以精确的方式执行逻辑功能。神经元彼此之间收集信息输入并加以处理，来决定是否发放输出信息，由此可以想象，神经元可能是活生生的逻辑门。他们推断，大脑可能是由神经元所搭建的and、or等逻辑门和其他逻辑元件构成，就像数字电路一样。我们不清楚麦卡洛奇和皮茨是否真的认为大脑是这样工作的，他们只是说，有可能是这样。而且从逻辑上讲，对于神经元的这种观点亦不无可能。从理论上讲，神经元可以实现数字功能。然而，没有人去追问，神经元在大脑中究竟以何种方式相连。尽管缺乏生物学上的证据，他们依然将这一理论作为“大脑只是另一种计算机”的证据。

还有一点值得我们注意的是，在20世纪上半叶，人工智能的理念得到了心理学中一个占主导地位的流派的支持，这一流派被称为行为主

义。行为主义者认为，大脑内部的运作机制是不可知的，他们将大脑称为“密闭的黑盒子”。但是人们可以观察和测量动物的环境和它们的行为——它们感受到了什么、做了什么，它们的输入信息和输出行为分别是什么。他们承认大脑具有反射机制，可用于训练动物形成条件反射，通过奖励和惩罚令它们学会新的行为方式。但是，除此以外，没有任何必要去研究大脑，尤其是那些混乱的主观感受，如饥饿、恐惧，或者理解某件事物的意义。不用说，这一研究理念最终在20世纪后半叶消亡了，然而人工智能领域却在很长一段时间内仍滞留于此。

二战结束后，电子数字计算机有了更广泛的应用，人工智能的先驱们纷纷卷起袖子开始了编程。实现语言之间的翻译？简单！就像破译密码一样，只需要将系统A的每个符号映射到系统B中的对应部分即可。处理视觉图像？似乎也不难。我们已经了解了处理图像旋转、缩放和位移的几何定理，而且能够轻松地将它们编成计算机算法代码，至此已经事半功倍了。人工智能专家们煞有介事地宣布，计算机的智能将很快赶上并最终超越人类。

具有讽刺意味的是，最有可能通过图灵测试的，是一个叫作Eliza的程序，它能够模仿精神分析师，将你的问题重新表述成新问题来反问你。例如，如果你输入“我和我的男朋友不说话了”，伊丽莎可能会回应：“跟我说说你的男朋友吧！”或者“是什么原因让你同你的男朋友不再说话了呢？”尽管它只是个被当成玩笑设计出来的无聊程序，但还是成功地骗过了不少人。也有一些正正经经设计出来的程序，比如积木世界（Blocks World），它模拟出一个包括许多不同颜色和形状积木的房间。你可以向程序提问，例如“在大红色方块上面有一个绿色金字塔形的积木吗？”或者向它发出指令，例如“请把蓝色方块移动到红色的小方块上面”，它能回答你的问题，或者按照你的要求工作。这一切都是模拟的——也确实管用。然而它只局限在完全虚拟的积木世界里，程序员们无法将其扩展到实际的应用中。

与此同时，人工智能技术领域的一连串表面上的成功和相关新闻给公众留下了深刻的印象。最初引起人们兴奋的是一个能够解决数学定理问题的智能程序。自柏拉图以来，多步演绎推理就被视为人类智慧的巅峰，因此起初看来，人工智能就像是中了头彩。然而结果证明，同“积木世界”一样，这个程序的能力也是有限的，它只能找出非常简单的定理，且都是一些已知定理。随后，一个名为“专家系统”的数据库引发了公众的巨大关注，它包含了能够回答人类用户所提问题的细节事实。比如，“医学专家系统”能够根据列出的症状来诊断患者的疾病。然而，它们再一次被证明作用有限，并且没有任何接近广义智能的表现。计算机程序还一度在棋盘类游戏中达到了专家水平，IBM的“深蓝”电脑对战国际象棋世界冠军加里·卡斯帕罗夫（Gary Kasparov）时的大获全胜曾经轰动一时。但是这些胜利毫无意义，因为“深蓝”电脑并非赢在比人类更聪明，而是赢在比人类快了几百万倍的运算速度上。“深蓝”没有直觉。人类象棋大师综观盘面，一眼就看得出棋盘上的有利和危险区域，而一台计算机对于这些重要信息没有天生的直觉，必须去试探更多的选择。除此之外，“深蓝”对于象棋的历史毫无概念，对自己的对手也一无所知。就像计算器能够做算术却不懂数学一样，“深蓝”能够下棋，但并不真正了解象棋。

在任何情况下，再怎么成功的人工智能程序也只是擅长于它们被设定处理的特定领域。它们无法概括，也不具有灵活性，甚至连它们的创造者也承认它们不会像人类一样思考。一些起初被认为容易解决的人工智能问题最后都无功而返。至今仍没有一台计算机能够在语言理解能力上超过3岁的幼儿，或在视觉能力上超过一只老鼠。

多年的努力换来的只是无法兑现的承诺和不被认可的成果，人工智能逐渐开始失去它的光环。一些科学家转向了其他领域。初创公司纷纷倒闭，资金支持也越来越少。就连编写一段执行感知、语言和行为等最基本功能的程序似乎都是不可能完成的任务。这种状况到今天也没有改变。正如我先前所说，尽管仍有人相信人工智能所面临的问题能够通过

运行更快的计算机来解决，但大多数科学家认为，以往所有的努力都存在问题。

我们不应该由此责难人工智能的先驱们。阿兰·图灵的确是伟大的。他们中的每个人都会告诉你，图灵机将要改变世界——它也确实改变了，但不是通过人工智能。

* * *

在申请MIT时期，我对于人工智能领域种种论断的怀疑进一步加深了。加州大学伯克利分校的著名哲学教授约翰·塞尔（John Searle）当时提出，计算机不是智能，也无法获得智能。为了证明这一点，在1980年，他想出了一个被称为“中文屋”的思维实验：

假设有间屋子，墙上开了一条缝。屋里的桌子旁边，坐着一个会说英语的人。他手头有一本很厚的说明手册，还有足够用的铅笔和草稿纸。通过翻阅手册，他能够根据用英文写成的说明，来处理、排序和比较汉字。重点是，手册中的指令同汉字的含义没有丝毫关系，它们只负责解决汉字应如何被复制、删除、重新排序和转录等问题。

屋外有人从墙上的缝塞进来一张纸。上面用中文写着一个故事和与之相关的问题。屋内的人对中文一窍不通，但他接过纸来，开始按照手册上的指令工作。这是一种生搬硬套的辛苦活，指令有时让他在纸上写下一些汉字，有时又让他移动或删除一些汉字。他按部就班地根据规则写写删删，直到指令告诉他工作已经完成为止。这时，他已经写出了一页新的汉字，这正是那些问题的答案，而他对此并不知情。按照指令，他需要将这页纸从缝中送出去。他照做了，心中却充满疑惑：这个乏味的游戏究竟是在做什么？

屋外，一个懂中文的女人读罢这页汉字后，表示答案完全正确——甚至还很有见地。如果你问她，这些答案是否出自于一个透彻地理解了

故事的聪明头脑？她一定会说是的。但她说得对吗？是谁理解了这个故事？当然肯定不是屋里的人——他完全不懂中文，对这个故事一无所知。但也不可能是那本手册吧——它只不过是一本安静地躺在纸堆里的书。那么，理解是从何处产生的呢？塞尔的回答是：根本没有“理解”这回事——有的只是无需动脑地翻书和写写画画而已。现在让我们转向问题的关键：“中文屋”与数字计算机何其相似！屋里的人就相当于是CPU，只会无意识地执行指令；手册相当于向CPU下达指令的软件程序；而那些纸就是内存。因此，一台通过产生相同的人类行为来模拟智能的计算机，无论设计得多么巧妙，也不会具有理解能力和智能。（塞尔曾明确表示，他不知道什么是智能——这句话的言外之意是，不管智能是什么，计算机肯定没有）。

这种说法令哲学家和人工智能专家之间产生了巨大分歧。它催生了数百篇夹杂着尖刻言辞的相互攻击的文章。人工智能的捍卫者们提出了许多论据来逐条反驳塞尔，例如，他们声称：虽然屋子的各组成部分都不懂中文，但如果将其视为一个整体来看，它还是懂的；屋里的人是懂中文的，只不过他没有意识到这一点。在我看来，塞尔的说法是对的。当我谨慎思考过“中文屋”实验的论证和计算机的工作原理之后，我并没有看到任何地方有“理解”的产生。这让我坚信，我们需要弄清什么是“理解”，并为它下一个定义。这个定义应当能够清楚地告诉我们，什么样的系统是智能的，什么样的不是；什么样的系统懂中文，什么样的不懂。而仅仅凭借系统的行为，是无法进行判断的。

人并不需要刻意去“做”任何事来理解一个故事。我可以安静地读一个故事，虽然没有任何外显的行为表明我清楚地理解了，但至少对我而言，这是个事实。另一方面，你无法从我安静的行为上看出我是否理解了故事，你甚至无从得知我是否懂得这个故事的写作语言。虽然过后你可以向我提问，但我对故事的理解发生在我阅读之时，而非回答之际。本书的其中一个主题便是：理解是无法通过外部行为来测量的，相反，它是对大脑如何形成记忆并利用这些记忆来作出预测的一个内部度量。

关于这一点，我们将在接下来的章节中谈到。“中文屋”、“深蓝”电脑和大多数计算机程序在这一点上有个相类似之处，就是它们并不理解自己在做的事情。而我们判断一个计算机是否智能，除了通过它的输出（即行为）外，并没有别的途径。

人工智能为自己辩护的最终论据是：理论上讲，计算机能够模拟整个大脑。一台计算机可以模拟所有的神经元和它们之间的连接，一旦它做到这一点，就意味着大脑“智能”和计算机的模拟“智能”之间不再有任何区别。虽然这在实际中不太可能，但我同意这一看法。遗憾的是，人工智能的研究者们并没有模拟大脑，因此他们的程序没有智能。而在理解大脑如何工作之前，也无法去模拟它。

* * *

被英特尔和麻省理工学院拒绝之后，我一时不知该何去何从。当你不知下一步怎么走时，最好的办法就是原地驻足，耐心等待转机的出现。于是我继续在计算机行业工作。我在波士顿过得心满意足，一直到1982年，我的妻子想要移居加州，我们便举家搬了过去（这也是让摩擦最小化的办法）。我在硅谷的一个初创公司Grid Systems找到了工作。Grid发明了第一台笔记本电脑，这台美妙的机器后来成为了纽约现代艺术博物馆的第一个计算机藏品。我先后在市场部和工程技术部工作过，直到后来我发明了一种高级编程语言，它被我命名为GridTask。随着这个发明连同我本人对Grid公司越来越重要，我的职业生涯也渐入佳境。

直到那时，我仍然没有办法将对大脑和智能机器的好奇抛诸脑后。我像着了魔一样地渴望去研究大脑。我报名了一个人体生理学的函授课程，开始自学（还好函授学校不会拒绝任何人！）。在学习了一定的生物学知识后，我决定申请一个生物学的研究生项目，打算从生物科学内部去研究智能。如果计算机科学界不需要大脑研究的理论家，或许生物学界会欢迎一个计算机科学家。那时还没有理论生物学这一科目，尤其是理论神经科学，因此对我的兴趣来说，生物物理学是最相近的领域。

我努力学习，参加了入学考试，准备了简历，恳请曾就职的公司写了多封推荐信，最后……乌拉！我终于被加州大学伯克利分校接收，成为一名生物物理学的全日制研究生。

我欣喜若狂地想，这下终于可以开始认真地研究大脑理论了！我从Grid公司辞了职，也没打算再重回计算机行业。当然，这意味着无限期地放弃我的高薪。我妻子那时正纠结于“是时候买房生娃过日子了”，而我却心甘情愿地成为一个不能养家糊口的人。这对于我们来说，绝对不是一个容易的选择。但对我个人来说，它却是最好的选择，我的妻子也因此支持了我。

在我离开公司之前，Grid的创始人约翰·艾伦比（John Ellenby）将我拉进他的办公室，对我说：“我知道你并没有打算再回到Grid公司或者计算机行业，但谁都不知道未来会发生什么，对不对？你与其完全退出，何不就把它当个休假呢？那样的话，如果一两年之内你返回来，就可以按照你离开时的标准，重新领薪水，继续做你的职位，拿你的股份。”这是一个非常友好的姿态，我于是接受了他的建议。但我的内心却明白，自己这次是要永远地告别计算机行业了。

第二章 神经网络

1986年1月，我开始在加州大学伯克利分校学习。我所做的第一件事，就是整理有关智能和大脑功能理论研究的历史。我阅读了上百篇由解剖学家、生理学家、哲学家、语言学家、计算机科学家和心理学家所著的论文。来自于不同领域的研究者们发表了大量关于思维和智慧的见解，各个领域都有专门的刊物和术语。然而，我发现这些见解既不一致，也不完整。当谈到智能时，语言学家总是会使用“句法”和“语义”等术语，在他们眼中，大脑和智能只同语言有关；视觉科学家习惯于谈论2D、2.5D和3D图像，大脑和智能对他们来说，只与视觉模式识别有关；计算机科学家们则津津乐道于由他们所提出的“模式”和“框架”等表征知识的新术语。没有人提及大脑的构造，也没有人关心这些理论在大脑中究竟如何实现。另一方面，解剖学家和神经生理学家撰写了大量有关大脑构造和神经元作用机理的论文，但对于建构大规模理论却退避三舍。毕竟，想要从各种研究方法以及随之而来的堆积如山的实验数据中寻找方向，实在是一件让人头痛的事。

就在此时，一种新的智能机器研究途径开始崭露头角，为人们带来了希望。虽然早在20世纪60年代后期，神经网络就已经开始以这样或那样的面目出现，但在当时，它同人工智能研究在投资份额和关注度方面存在着激烈的竞争。人工智能就像是一只体重800磅的大猩猩，将神经网络研究压制得无法抬头。神经网络的研究者在许多年间一直被列于投资方的黑名单上，只有少数人还在继续关注他们。直到20世纪80年代中期，这一领域才终于得以重见天日。我们很难确切地知道，神经网络为何突然变成了热点，但人工智能的节节失败无疑是其中的因素之一。人们在寻找人工智能的替代品，而最终在神经网络领域看到了希望。

相对于人工智能的方法，神经网络算得上一个真正的进步，因为它的架构建立在真正的神经系统之上，尽管根基尚浅。与计算机程序员不同，神经网络的研究人员（也被称为联结主义者）的兴趣在于了解，如果将一群神经元聚在一起，它们会表现出何种行为。大脑由神经元组成，因此构成了一个神经网络，这是铁一样的事实。联结主义者们希望通过研究神经元之间的相互作用，弄清智能那难以捉摸的特性；他们还希望通过复制神经元群之间的连接，解决那些令人工智能一筹莫展的问题。神经网络与计算机的不同之处在于，它没有CPU，也不需要中央存储。整个网络中的知识和记忆都分散在它的连接上——就像真正的大脑一样。

从表面上看，神经网络似乎非常符合我的兴趣。但很快我对这一领域的希望就又幻灭了。那时，我已经形成了一个自己的看法，即：对于大脑的理解，有3个标准是必不可少的。第一个标准是，对于大脑功能的理解，必须要考虑时间因素。真正的大脑始终在处理快速变化的信息流。在进出大脑的信息流中，没有什么是静止不动的。

第二个标准是，反馈的重要性。神经解剖学家一早就发现，大脑中充满了反馈连接。比如说，在新大脑皮层和丘脑之间连接的神经回路中，反馈连接（信息传递朝着输入的方向）的数目要比前馈连接多出将近10倍！也就是说，对于每一束向大脑皮层传递信息的神经纤维，都对应着10束向感觉器官传递信息的神经纤维。大脑皮层中的神经连接也绝大多数具有反馈功能。虽然反馈的确切作用尚无人知晓，但从已发表的研究报告中可以看出，它无处不在。据此我认为，反馈一定非常重要。

第三个标准是，任何理论或有关大脑的模型，都应该能够解释大脑的物理结构。新皮层并不是一个简单的构造，大家在后面的章节中将会看到，它有着不断重复的层级结构。任何不同于这一构造的神经网络，必定无法像大脑一样工作。

然而，神经网络刚一亮相，就定位于一些极为简单的模型上。这些

模型对于上述三个标准无一满足。绝大多数神经网络都是由相互连接的三排神经元组成。第一排神经元接受某种模式（输入），接着这些输入神经元同下一排神经元相连，我们称这些为“隐藏单元”。“隐藏单元”再与最后一排神经元（输出单元）相连。神经元之间的连接强度有强有弱，按照连接强弱的不同，一个神经元的活动可能会促进另一个神经元的活动，也可能会减弱第三个神经元的活动。神经网络就是通过改变这种连接强度，来学习如何将输入模式映射到输出模式上。

这些简单的神经网络只能用来处理静态模式，不涉及反馈，同大脑也没有任何相似之处。有一种最常见的神经网络，被称为“反向传播（back propagation）”网络，它能将一个错误从输出单元向输入单元传播来进行学习。你可能会认为这是反馈的一个形式，而事实上它不是。这种对错误的反向传送只发生在学习阶段。当神经网络经过训练，工作状态正常时，信息便只会向一个方向传送。在输出到输入的方向上，并无反馈发生。除此之外，这些模型中没有时间：一个静态输入模式被转化为一个静态的输出模式，紧接着又出现另一个输入模式。在这些网络中，哪怕对于刚刚发生的事情也不留存任何历史记录。最后，与大脑的复杂性及其层级结构相比，神经网络的构造显得太小儿科了。

我本以为神经网络领域会飞快地往更加仿真的网络发展，但它并没有。由于简单的神经网络已经能够做出一些有趣的事情，因此许多年后，研究还一直停留在这个层面。这种新鲜有趣的工具，一夜之间让成千上万的科学家、工程师和学生获得了资助、博士学位，发表了著作。利用神经网络进行股票市场预测、处理贷款申请、核对签名以及执行上百种其他模式分类应用的公司，也如雨后春笋般纷纷成立。尽管神经网络创建者的意图可能在于更为广泛的应用，然而当时在该领域居于主导地位的人们，对理解大脑如何工作以及什么是智能等问题，没有丝毫兴趣。

大众媒体对神经网络与智能之间的差别也不甚明白。报纸、杂志和

电视科学节目将神经网络介绍为“像大脑一样的”或是“以大脑工作原理为蓝本”。与处处需要编程的人工智能不同，神经网络通过事例进行学习，这让它多少看起来更智能一些。NetTalk即为其中的一个突出代表，它能够学着将字母顺序同读音一一匹配。由于这个神经网络是用印刷文本来训练的，因此它乍听起来就是用计算机的声音在朗读单词。不难想象，用不了多久，神经网络就可以同人类对话了。在全国新闻中，NetTalk被错误地介绍为一种能够学习阅读的机器。它虽然是神经网络的一个精彩展示，但所做的事情仍属于微不足道。它不会阅读，不能理解，且没有什么实用价值。它所做的只是将字母组合同预定的声音模式相匹配。

请允许我用一个类比来说明神经网络与真正的大脑之间差得有多远。想象我们要研究的不是大脑的原理，而是一台数字计算机。经过多年研究后，我们发现计算机中的一切都是由晶体管构成的。亿万的晶体管以精确而又复杂的方式连接在一起。然而我们仍然不明白计算机是如何工作的，也不明白这些晶体管为什么要以这种方式相连。于是某一天，我们决定将几个晶体管连接起来看个究竟。结果我们发现，瞧，将区区三个晶体管以某种方式连接在一起，就构成了一个放大器，一端输入的信号在另一端就会被放大。（收音机和电视机里的放大器就是用晶体管以这种方式制成的。）这是一个重大的发现，一夜之间，使用晶体管放大器制造收音机、电视机和其他电子设备的新工业产生了。这固然是好事，但它还是没能告诉我们计算机是如何工作的。尽管放大器和计算机都是由晶体管构成的，但它们之间几乎再没有别的共同之处。同理，尽管真正的大脑同三排的神经网络都是由神经元构成，它们也几乎完全不同。

我在1987年夏天遇到的一件事，又在我对神经网络本来就不太高的兴趣上泼了盆凉水。当时我参加了一个神经网络的会议，期间观看了一家名为Nestor的公司的展示。Nestor推出了一种在平板电脑上识别手写文字的神经网络应用，要价100万美元。这引起了我的注意。虽然Nestor

大力鼓吹它的神经网络算法有多么复杂精妙，甚至将其吹捧为另一个重大性突破，但我却觉得手写识别问题其实可以通过更为简单、传统的方法解决。那天我回到家里，反复思考这个问题。两天后，我设计出了一款速度更快、体积更小、使用更灵活的手写识别器。我的解决方案里并没有使用到神经网络，其工作原理也同大脑完全不同。尽管那次会议引发了我对设计带有触控笔界面的电脑的兴趣（并最终成就了10年后的Palm Pilot掌上电脑），它同样也使我更加确信，神经网络相对传统方法而言，并无太大的改善。我设计的手写识别器最后成为了Graffiti文本输入系统的基础，被广泛应用于第一代Palm产品上。我想Nestor在这场商业竞争中应该是被淘汰了。

简单的神经网络走到了尽头。它们的大多数功能都能被其他方法轻易取代，最终媒体的关注热情也逐渐消散。但至少，神经网络的研究者们并没有宣称他们的模型是智能的，毕竟它们只是些极其简单的网络，功能上也没有超越人工智能。我在此并不想给大家留下一种印象，认为所有的神经网络都只有简单的三层变化。一些研究人员仍在继续研究设计不同的神经网络。如今，这个名词被用来描述一系列不同模型的集合，其中一些在生物学看来是精确的，另一些则不是。但它们几乎都没有抓住新皮层的总体功能和结构。

在我看来，大多数神经网络的最根本缺陷在于——这也是它与人工智能共有的特点——太注重行为。这是一个致命的负担。无论他们将这些行为称为“答案”、“模式”，还是“输出”，人工智能和神经网络研究者都假定智能存在于一个程序或神经网络处理输入信息之后而产生的行为中。计算机程序或神经网络最重要的属性就在于它是否能给出正确的、令人满意的输出。就像阿兰·图灵所给出的启示，智能等同于行为。

然而，智能并不单是指表现出智能的动作或行为。行为是智能的一种表现，但它既不是智能的核心特征，也不是智能的基本定义。片刻的思考就可以证明这一点：即使躺在黑暗中什么都不做，只是思考和理

解，你也是智能的。忽略头脑中的活动而只关注于行为，对理解智能和建造智能机器造成了极大的障碍。

* * *

在进一步探索智能的新定义之前，我想先介绍另一种与真正大脑的工作原理更为接近的联结主义方法。问题是，似乎没有人认识到这项研究的重要性。

就在神经网络大出风头之时，一小部分研究神经网络理论的学者从主流领域中分离出来，构建了一种不以行为为中心的网络，称之为“自-联想”记忆网络。它同样由相互连接的简单神经元构成，这些神经元在达到一定刺激阈值时会激活。然而它们之间的连接方式与一般的神经网络不同，其中使用了大量的反馈。与只能正向传输信息的神经网络不同，自-联想记忆与反向传播网络类似，能将每个神经元的输出传回给输入——就像自己给自己拨电话。这种反馈回路造成了一些有趣的特点。当一种活动的模式被加予人造神经元时，它们会对这种模式形成记忆，这种网络将外界活动模式同它自身关联在一起，因此被称为“自-联想”记忆。

初看起来，这种回路所导致的结果似乎很荒谬。想要检索一个被存储于这种记忆中的模式，你必须先提供这个模式。这就好比你去杂货店买香蕉，当店主问你如何付款时，你说用香蕉。你可能会问：“这样的设计有什么好处呢？”然而，自-联想记忆所拥有的一些重要特征，在大脑中亦有体现。

其中最重要的一个特征是：如果想要检索某个模式，你不必事先拥有这个模式的全部，只要有其中的一部分甚至一个乱作一团的样子就可以。即使从一个混乱的版本开始，自-联想记忆也可以检索到最初存储时的正确模式。这就好比拿着吃剩的半把褐色香蕉去杂货店换回了一整把绿色香蕉一样。或是你拿着残破得无法辨认的钞票来到银行，柜台职

员对你说：“我看得出这是一张破损的百元大钞，来把它给我，我给你换一张崭新的。”

第二个特征是，与大多数其他的神经网络不同，自-联想记忆可被设计用来存储模式序列，或称为时序模式。这一功能可以通过在反馈中加入延时来实现。有了这个延时，你便可以向该网络呈现一个模式序列，类似于一段旋律，自－联想记忆就可以记住它。当我输入“一闪一闪亮晶晶”的前几个音符时，自-联想记忆马上就可以返回给我整首曲子。当输入序列的一部分时，该记忆便能够回忆起其余的部分。我们将会看到，这同人们学习几乎所有模式序列时的方式如出一辙。我认为，大脑就是使用与自-联想记忆相似的回路来实现这种学习的。

自-联想记忆提示了反馈和随时间变化的输入的潜在重要性。遗憾的是，绝大多数的人工智能、神经网络和认知科学家们都忽视了这两者。

神经科学家作为一个整体，表现也差强人意。他们当然也了解反馈（就是他们最先发现的），但他们当中大多数人并没有提出什么理论（除了模糊地谈到“阶段”和“调制”之外）来解释大脑为何需要这么多的反馈。在他们对于大脑总体功能的看法中，时间所扮演的角色也是可有可无的。他们往往以图表的方式说明事件发生在大脑的哪个部位，而不关心神经元激活模式在何时或以何种方式产生时序上的相互作用。出现这种偏见的原因，一部分来源于目前实验技术的限制。20世纪90年代，这一被称为“大脑研究的十年”中，最受欢迎的技术之一是功能性核磁共振成像。功能性成像仪可以拍下人脑活动的照片。然而，它无法记录快速的变化。因此，科学家们要求被试们一遍又一遍地集中注意处理单一任务，就像在拍光学照片时让他们保持不动一样，只不过功能性成像是对大脑精神活动的拍摄。结果，我们得到了大量的数据用来说明某种特定任务激活了大脑的哪个部位，但用于说明随时间变化的真实信息输入如何流经大脑的数据，仍旧少得可怜。功能性成像技术使我们能够观察

到特定时刻的事件发生的脑区，却无法记录下大脑活动是如何随时间变化的。科学家们固然想要收集这方面的数据，但苦于没有合适的技术支持。也是因为这样，许多主流的认知神经科学家仍然相信从输入到输出的谬论——给出固定的输入后，等着看有怎样的输出。大脑皮层的连线图往往以流程图来表示：信息从初级感觉区——视觉、听觉和触觉进入的区域，向上传至较高的分析、规划和运动区，然后再将指令向下传达至肌肉。你先有感觉，然后才有动作。

我并不是说所有人都忽略了时间和反馈。这一领域如此广阔，几乎任何一种想法都会有追随者。近年来，已经有越来越多的人意识到反馈、时间和预测的重要性。然而笼罩在人工智能和经典神经网络周围的光环，黯淡了其他的研究方法，使得它们未能得到充分的认识。

* * *

无论外行还是专家，都认为智能应该由行为定义。这一点也不难理解。至少近几个世纪以来，人们一直不断地将大脑的功能同钟表、水泵系统、蒸汽机，再到后来的计算机联系在一起。近几十年来的科幻小说中一直充斥着人工智能的思想，从艾萨克·阿西莫夫（Isaac Asimov）的机器人三大定律，到《星球大战》中的机器人C3PO，智能机器能够帮人做事的观念在我们的想象中根深蒂固。所有的机器——无论是由人类建造的还是想象中的——都被设计为能够帮我们做事情。我们没有会思考的机器，只有能干活的机器。甚至当我们在观察我们的人类同胞时，所关注的也是外在行为，而非他们隐藏于内心的想法。因此，从直觉上看来，一个智能系统的衡量标准显然应该是智能行为。

然而，回望整个科学发展的历史，我们会看到，直觉往往是发现真理的最大阻碍。科学的框架之所以难以发现，并不是因为它们复杂，而是因为直觉上的错误假设让我们看不见正确答案。哥白尼（Copernicus，1473—1543）之前的天文学家错误地认为地球是固定不动的，且处于宇宙的中心，就是因为它让人感觉平稳，而且看起来像是

宇宙中心点。直觉告诉我们，星星是一个巨大的旋转球面的一部分，而我们在其球心位置。如果有人说，地球像陀螺一样旋转，它的表面每小时移动近千公里，而且整个地球还在太空中疾驰（更别提那些星星离我们有上亿万公里的距离了），大家一定会认为他疯了。然而，这一切都被证明是正确的。其实它们都很好理解，无奈却不符合直觉。

在达尔文（1809—1882）之前，物种的形式在人们看来显然是固定的。鳄鱼同蜂鸟不可能相配，它们不仅不同，甚至相互对立。物种进化的思想不仅与宗教教义相悖，而且违背了人们的常识。进化意味着你同每一个生活在这个星球上的生物都拥有共同的祖先，这些生物包括蠕虫和厨房里摆放的开花植物。我们现在知道这是真的，而直觉却相反。

之所以提到这些著名的例子，是因为我相信，对于智能机器的追求同样受到了直觉假设的阻碍，因此才使得我们进展缓慢。当你问自己“智能系统可以做什么”时，直觉就会理所当然地从行为上去思考。你可能会说，我们的确是通过说话、写字和行为展现出人类智能的，难道不是这样吗？当然是，但这只是一个方面。智能在你头脑中发生，行为只是它的一部分。这一点虽然并不直观，但理解起来应该也不算难。

* * *

1986年的春天，我日复一日地坐在办公桌前，阅读科学论文，整理智能研究的历史资料，并时刻关注着人工智能和神经网络领域的不断变化。我渐渐发现自己被细节淹没了。虽说有无穷无尽的知识供我研究和阅读，但对于整个大脑究竟是如何工作的，我仍然没有得到任何清楚的答案。我甚至不清楚它究竟做了什么。这要归咎于神经科学本身就充满了细节，直到现在也仍然如此。每年都有数以千计的研究报告发表出来，但研究者们往往只是堆砌新的细节，而不是将它们组织起来。直到现在，仍然没有一个整体的理论和框架能够解释大脑在做什么以及如何做。

从那时起我便开始思考，这个问题的解决方法应该是什么样的？它会因为大脑本身的复杂而极尽复杂吗？是否要用到写满100页纸的数学公式才能描述出大脑的工作原理呢？我们是否需要绘制出成百上千的独立线路图才能理解任何有用的信息？我并不这么认为。历史证明，科学问题的最佳解决方案往往是简洁而优雅的。虽然细节可能令人生畏，通往最后理论的道路可能崎岖，但最终的概念框架通常很简单。

如果没有核心解释作为引导，神经科学家将无法把已经收集到的所有细节统一起来，形成连贯的画面。大脑由无数神经元缠连而成，其复杂程度令人咋舌。初看起来，它就像一个装满了煮熟意大利面的体育场。也可以称之为电工的噩梦。但如果仔细观察，我们就会看到，大脑并不是杂乱无章的。它有大量的组织和结构，但这些组织和结构对我们来说过于复杂，以至于我们无法凭直觉去理解它们的整体运作，毕竟这不像将一个打碎的花瓶碎片重新拼起来那样简单。这种努力上的失败并非是因为没有足够的或正确的数据，我们所需要的，是转换思考的角度。在正确的框架下思考，细节才能有意义和可操控。看看下面这个虚构的比喻，你就会明白我的意思。

幻想一下：距今一千年以后，人类已经灭绝。来自外星高级文明的探险者登陆地球，想要弄清楚人类是如何生活的。他们尤其困惑于我们的道路网络：这些奇形怪状的复杂结构都是干什么用的？他们通过卫星和地面探查建立索引，像一丝不苟的考古学家那样，记录下每一块沥青碎片的位置，每一个被风吹日蚀的路标，以及所有可以找到的细节。渐渐地，他们注意到，有些道路网络与其他的不同：在有些地方它们狭窄崎岖，看上去毫无规划；有些地方它们又形成了规则完美的网格，延伸了一段之后变得宽阔起来，绵延数百英里穿越沙漠。他们收集了堆积如山的细节，但这些细节无法告诉他们任何事情。他们只好继续收集更多的细节，希望能找到新的数据解释这一切。在很长一段时间里，他们被这个问题困住了。

直到有一天，其中一个家伙说：“有了！我想我明白了……这些生物不像我们一样能瞬间移动。他们必须使用一种设计巧妙的移动平台，从一个地方移动到另一个地方。从这个基本的认识出发，许多细节变得明朗起来。弯弯曲曲的小街道网络是早期运输工具还很慢的时候修建的，而又宽又长的道路是用于长途高速运输的。这也最终解释了为什么道路上画有不同的数字标志。外星科学家们由此开始推断哪里是居民区，哪里是工业区，甚至商贸需求和交通运输的基础设施之间相互作用的方式，等等。他们之前整理出的许多细节变得不再相关，而只是出于历史的偶然或当地地理状况的要求。然而此时，同样的一堆原始数据已经不再令他们困惑了。

我相信，同样类型的突破也会帮助我们了解有关大脑细节的真相。

* * *

令人遗憾的是，并非每个人都相信我们能够了解大脑是如何工作的。相当多的人——其中还包括一些神经学家——认为，从某种程度上来说，大脑和智能是无法解释的。还有一些人认为，即便我们理解了它们，也不可能建造出有着同样工作原理的机器，因为智能必须要以人体、神经元，甚至某些高深莫测的新的物理定律作为基础。每当我听到这些争论，就会想起过去的老学究，他们反对研究天空，也反对通过解剖尸体研究人体的工作原理。

“不要费劲研究那些，没有意义，因为即使你弄清了它是如何工作的，对我们来说也没什么用。”类似这样的争论将我们引向哲学的一个分支——功能主义，在短暂的思维研究史上，它是距离我们最近的一站。

根据功能主义的看法，智能和心灵是纯粹的组织特性，它们与你的身体由什么材料构成并没有本质上的关系。心灵可以存在于任何系统之中，只要该系统的组成部分之间有正确的因果关系，而这些部分可以是

神经元、硅芯片或者其他任何东西。很显然，这种观点是任何想要建造智能机器的人的标准说辞。试想一下：如果用一个小盐瓶来代替棋盘上丢失的马，这盘棋会因此而丧失哪怕一丝真实性吗？显然不会。小盐瓶在功能上等同于一个真正的马，因为它按照马的规则在棋盘上移动和与其他棋子配合，所以这盘棋仍是一局真正的比赛，而不仅仅是一个模拟游戏。再考虑另外一种情况：如果我用光标将这句话中的每个字符都删掉，然后再重打一遍，难道这句话就会有所不同吗？还有一个同我们切身相关的例子：每隔几年，组成你身体的大部分原子就会更新换代。尽管如此，从各种重要角度来看，你还是你。如果一个原子在你的分子构成中的功能性作用与其他原子相同，它们就是等价的。这种说法同样适用于大脑：如果一个疯狂的科学家将你的每一个神经元都用具有等价功能的微机械复制品代替，最后产生出的那个你所感受到的真实自我，同原来没有任何差别。

根据这个原理，一个使用与大脑相同的智能构造的人工大脑，应当和大脑一样聪明，并且不只是“人工的”智能，而是“真正的”智能。

人工智能的支持者、联结主义者和我，大家都是功能主义者，因为我们都相信大脑的智能并没有什么特别的神秘之处，总有一天我们会用某种方式建造出智能机器。然而，我们在对功能主义的解释上存在着不同。虽然我已经表述过人工智能和联结主义范式的主要失败之处，即输入——输出谬误，但我觉得还是有必要再说一说，为什么我们还不具备设计智能机器的能力。在我看来，人工智能的支持者们所采取的是一种弄巧成拙的强硬激进路线，而联结主义者则主要是太保守了。

人工智能研究人员会问：“为什么我们这些工程师要被进化中偶然发现的解决方案所束缚呢？”在原则上，他们的话也有点道理。像大脑和基因组这样的生物系统，都是出了名的粗陋不堪，并不够优雅。对此一个常见的比喻是鲁贝戈德堡机器（Rube Goldberg machine），它以大萧条时期的漫画家鲁贝戈德堡命名，指他的漫画中那些用来完成微不足

道的任务、但却过分复杂的装置。软件设计师们也有一个相应的术语——“组装件”（kludge），指那些由于缺乏远见而最终充满了累赘和无用的复杂性，往往连程序编写者自己也难以理解的程序。人工智能研究人员担心，大脑也是一个同样的烂摊子，一个进化了几亿年的“组装件”，充斥着低效能以及进化中的“遗留代码”。如果真是这样，他们说，为什么不把这个令人失望的混乱体完全丢弃，重新开始呢？

许多哲学家和认知心理学家都赞同这种观点。他们喜欢将思维比喻为大脑中运行的软件，将大脑比喻为有机的计算机硬件。在计算机中，硬件和软件是完全不同的级别。同一个软件程序可以在任何一台通用图灵机上运行，比如你可以在个人电脑、苹果机或克雷（Cray）超级计算机上运行WordPerfect，虽然这三个系统有着不同的硬件配置。而如果你要学习WordPerfect的话，硬件系统则一点忙都帮不上。依此类推，大脑无法告诉我们什么是思维。

人工智能的拥护者们还喜欢搬出一些历史实例来支持他们的观点。对这些例子中的问题，工程学给出的解决方案从根本上不同于大自然的版本。比如，我们是如何成功制造出飞行器的？是通过模仿有翼类动物扇动翅膀的动作吗？显然不是。我们用的是固定机翼和螺旋桨，后来又使用了喷气发动机。它虽然不是大自然的解决方案，但却很有用——甚至比上下扇动的翅膀更出色。

同样，我们并没有通过模仿猎豹的四条腿，而是通过车轮的发明造出了速度超过猎豹的陆地运输工具。轮子的使用是在平坦陆地上移动的好主意，我们不能因为进化中没有采用这种方式，而否定它是一个可行的好办法。一些研究思维的哲学家对于“认知轮子”的比喻一见倾心，也就是说，人工智能对一些问题的解决办法，虽然完全不同于大脑，但也同样出色。换句话说，一个工作方式有限但有效的程序，如果在某个任务上的表现堪比（或者超越）人类，那么它就与我们大脑的能力不相上下。

我认为正是这种“只问结果，不问手段”的功能主义解释，将人工智能研究者们引入了歧途。正如塞尔在“中文屋”实验中所揭示的那样，行为上的等同是不够的。既然智力是大脑的内部属性，那么我们就必须探究大脑的内部来了解什么是智能。在对大脑，尤其是新大脑皮层的研究中，我们需要仔细甄别哪些细节是进化过程中多余的“冻结的偶然事件”（frozen accident）；毫无疑问，在许多重要功能中也会混杂着鲁贝戈德堡式的无用处理。但正如我们很快将要看到的，在这些神经回路所体现出的强大能力中，蕴藏着一种优雅的原理，正等待着我们去发现，它超越了世上任何最先进的计算机。

联结主义者凭直觉感到，大脑不同于计算机，而它的秘密就在于神经元相连时的作用机理。这是一个好的开端，但该领域至此就停滞不前了。虽然曾有上千人致力于三层神经网络的研究，至今仍有一些人在努力，但对于真实皮层网络的研究却仍旧是凤毛麟角。

半个世纪以来，我们调用了人类全部的聪明才智，试图将智能赋予计算机。在这个过程中，我们发明了文字处理器、数据库、视频游戏、互联网、手机和逼真的电脑动画恐龙，但智能机器仍然杳无踪影。要想取得成功，我们必须努力模仿大自然的智能引擎——新皮层，将智能从大脑的内部提取出来。除此之外，别无他途。

第三章 人脑

是什么让大脑与人工智能和神经网络的程序设定如此不同？它究竟有何独特之处？这种独特又为何如此重要呢？在接下来的几章中，我们将看到，这些问题的答案同大脑的构造有很大的关系，它将告诉我们大脑是如何工作的，以及为什么它与计算机完全不同。

让我们先来看看这个器官。假设桌上有一个大脑，我们正在一起解剖它。你首先会注意到的是，大脑的外表非常均匀——颜色是粉灰色，样子有点像表面光滑的花椰菜，上面布满了脊突和沟壑，被称作“脑回”（gyri）和“脑沟”（sulci）。外面摸起来软软黏黏的部分就是新大脑皮层，它是一层薄薄的神经组织，包裹着大部分旧脑。我们将把重点集中在新皮层上，因为几乎所有被我们视为智能的能力，如感知、语言、想象力、数学、艺术、音乐以及计划，都发生在这里。此刻，正是你的新大脑皮层在阅读这本书。

写到这儿，我要向大家承认，我是一名新皮层至上主义者。这一点，可能会遭遇一些人的反对。因此在深入讨论前，请允许我用一分钟的时间，说一说我的理由。大脑的每个部分都有专门的科学家研究团体研究它，如果有人说，通过解读新皮层就可以彻底了解智能的本质，那肯定会激起一些研究团体的反对。他们会说：“如果不先了解大脑某某区域的话，你是不大可能了解新皮层的，因为它们之间的联系非常紧密，并且你需要这个区域来完成某某工作。”对此说法我并没有异议。诚然，大脑包括许多部分，其中绝大多数对人类来说都至关重要。（奇怪的是，小脑是个例外。虽然它包含了最多的神经细胞，但如果你天生就没有小脑，或者小脑损伤的话，仍然可以过正常的生活。但大多数其他脑区却不行，它们是维持基本生活和感知所必需的。）

对于那些反对的声音，我想说的是，我对造人没有兴趣。我只是想要理解智能，并建造智能机器。拥有人类特质和拥有智能是两码事。智能机器不需要性驱力、饥饿感、脉搏、肌肉、情绪，或一个像人一样的身体。人不仅仅是智能机器。作为大自然的创造物，我们身上既有着生物体必需的特质，也不乏由亿万年的进化留下的冗余之物。如果你想建造像人类一样的智能机器，使它在各个方面都能通过图灵测试，那么你很可能将需要重新造出许多作为一个人所必需的其他东西。然而正如我们将要看到的，如果是建造不完全等同于人类的智能机器，我们只需专注于大脑中同智能紧密相关的部分即可。

对于那些被此观点冒犯的人，我想说明的是，我同意大脑的其他结构，如脑干（brain stem）、基底神经节（basal ganglia）和杏仁核（amygdala）等，确实对新皮层的功能运作至关重要。这一点毫无疑问。但我希望你能理解，智能的一切重要方面都发生在新皮层。另外两个脑区——丘脑（thalamus）和海马体（hippocampus）——也发挥了重要的作用。对于这两个结构，我们将在后文中讨论。从长远来看，我们需要了解所有脑区的功能和作用。但我相信，在一个完善的新皮层功能理论的背景下，这一切将能得到最好的解决。这就是我的两点看法。现在，让我们重新回到新大脑皮层上来。你也可以简称它为皮层（cortex）。

拿出6张名片或者扑克牌，并将它们叠在一起。（真的这样去做会比你光凭想象有帮助得多。）现在你手中正拿着一个大脑皮层的模型。约2毫米厚的6张名片可以帮助你感受大脑皮层有多薄，它就像你手中的这叠卡片一样，约2毫米厚，有6层，每一层的厚度相当于一张卡片。

如果将人类的大脑皮层完全展开，其面积大约相当于一张大的晚餐餐巾。其他哺乳类动物的皮层则要小得多：老鼠的皮层相当于一张邮票；猴子的则相当于一个商业信函的信封。但无论大小，大部分皮层都像你手中的名片一样，分为6层。人类之所以更聪明，是因为我们的大

脑皮层相对于我们的体型来说，覆盖了更大的区域，而不是因为它更厚或是含有某类特殊的“智能”细胞。它的大小令人印象深刻，因为它环绕并包裹了大脑的大部分。为了适应这个大脑袋，大自然不得不改变人体的解剖结构，于是人类女性进化出了一个能够分娩出大头婴儿的宽骨盆。一些古人类学者认为，这一特征是同双腿直立行走能力共同进化出来的。然而这仍不够，因此进化又将大脑皮层折叠起来，塞进我们的颅骨，就像把一张揉皱的纸塞进白兰地酒杯中。

你的大脑皮层中遍布神经细胞，也叫作神经元，它们排列得非常紧密，没有人能说出它们的准确数目。如果在你手中那叠卡片上画一个边长一毫米的小方块，就可以标记出大约10万个神经元的所在位置。想在这样微小的空间里数出这个确切的数字，几乎是不可能的。尽管如此，一些解剖学家估计，人类大脑皮层中包含大约300亿个神经元。但就算这个数字再大或者再小很多，也没有人会感到惊讶。

这300亿个神经元造就了你。他们包含了你几乎所有的记忆、知识、技能，还有你所积累的丰富生活经验。尽管已经研究了25年大脑，我仍然觉得这一事实令人震撼：就是这样一层薄薄的细胞，让我们有能力去看、去感知，并形成对世界的看法，这真是不可思议。夏日的温暖和对美好世界的憧憬，都是这些细胞的杰作。曾在《科学美国人》上发文的弗朗西斯·克里克（Francis Crick），很多年后又写了一本书关于大脑的书，书名叫作《惊人的假说》（*The Astonishing Hypothesis*）。他所谓的惊人的假说是指，思想仅仅是大脑细胞的产物——没有魔法，没有特调酱汁，只有神经元和信息流之舞。我希望你们能够体会到这有多么不可思议。细胞群同意识体验之间似乎存在着巨大的哲学鸿沟，但思想和大脑是一体的。克里克那时称其为假说，只是政治正确的说法。因为思想由大脑细胞所创造，这是一个事实，而不是假说。我们需要弄清的是，这300亿个细胞究竟做了什么、是如何做的。所幸构成大脑皮层的细胞并不是混乱无序的一团，我们可以更深入地了解它的结构，进而探索它是如何创造出人类思想的。

* * *

让我们回到解剖台，再继续观察大脑。对于肉眼来说，大脑皮层上几乎没有任何可以作为标识的地方。当然真要找的话也有那么几个，比如将大脑分为左右两个半球的纵向裂，还有一个将大脑分为前后两个区域的回间沟。但除此以外，无论从左至右还是从后向前，目光所及之处，就都是看起来毫无二致的错综复杂的表面了。并没有明显的边界线或是颜色标记能够帮你区分处理不同感官信息或不同类型思想的脑区。

然而，人们很早以前就知道，大脑中存在分界。甚至早在神经科学家能够识别同大脑皮层回路相关的东西之前，人们就已经意识到，某些心理功能是定位到特定区域上的。如果中风损伤了一个人的右顶叶，他可能将会失去对左侧身体或左侧空间的感知能力，甚至是想象能力。相比之下，如果中风发生在他左侧额叶的布洛卡区（Broca’s area），就会影响他运用语法规则的能力，而他的词汇量和理解词汇含义的能力不会改变。如果中风发生在一个被称为梭状回（fusiform gyrus）的区域，就会影响他识别面孔的能力——他将无法认出自己的母亲和孩子们，甚至无法认出照片上自己的脸。这些奇特的功能障碍，让早期的神经科学家们意识到，大脑皮层由许多功能部位或功能区组成——这两个术语是等价的。

在过去的一个世纪里，我们对大脑功能区已经有了相当的了解，但仍然有许多问题亟待解答。每一个功能区都是半独立的，似乎专门负责知觉或思维的特定方面。它们东一块西一块地分布于大脑皮层上，形成“百衲被”一样毫无规则的图案，然而在人与人之间的差别微乎其微。这些功能极少有清晰的界限。在功能上，它们是按照一种分支的层级模式排列的。

鉴于层级概念的重要性，我想多花点时间来仔细界定它。这一概念将贯穿全书。在一个层级系统中，某些元素在抽象意义上“高”或“低”于其他元素。比如在商业层级中，中层经理这一职位居于收发员之上、副

总裁之下。这同物理上的“高、低”概念没有关系，即使所在的楼层比收发员低，经理的级别仍然是高于收发员的。我之所以要强调这一点，是为了让大家明白，当我说一个功能区比另一个功能区更高或更低时的含义。这种高低同它们在大脑中所处的物理位置无关。所有的功能区都分布在同一个复杂的大脑皮层上。一个区域比另一个区域“高”或“低”的关键，取决于它们之间的连接方式。在大脑皮层上，低级区域通过特定的神经连接模式将信息上传给高级区域，而高级区域用另一种方式向下发送反馈给低级区域。层级结构不同分支下的区域之间也存在着连接，就像一个地区的中层经理与另一地区的同级别经理之间存在交流一样。科学家丹尼尔·福尔曼（Daniel Felleman）和大卫·范·埃森（David van Essen）绘制出了详细的猴子大脑皮层图。图中显示，许许多多的功能区在复杂的层级结构中彼此相连。由此我们可以假设，人类大脑皮层也有着类似的层级结构。

最低级的功能区——初级感觉区，是感觉信息最先到达的皮层区域。这一区域在最原始和最基本的层面上处理信息。例如，视觉信息通过简称为V1的初级视觉区进入大脑皮层。V1负责对低级视觉特征的检测，如一小段边缘线、小幅运动、双眼视差（立体视觉）以及基本色彩和对比度信息。随后，V1将这些信息上传给如V2、V4、IT（我们将在后面详细介绍）以及一系列其他区域。它们中的每一个都负责信息中某些更专门和抽象的方面。例如，V4中的神经元会对中等复杂的模式，如红色或蓝色的星形产生反应。另一个被称为MT区的区域则专门负责检测物体的运动。位于视觉皮层更高层级上的是一些表征你对各种物体（如面孔、动物、工具、身体部位等）的视觉记忆的区域。

你的其他感觉也对应着类似的层级结构。大脑皮层上有一个被称为A1的听觉区和更高级的听觉层级结构，以及一个被称为S1的初级体感（躯体感觉）区和更高级的躯体感觉层级结构。最后，这些感觉信息会进入到“联合区”，这个区域一般接收来自多个感官的输入。例如，某些皮层区域同时接收来自视觉和触觉的输入。正是因为有了这些联合区，

你才能将看到苍蝇趴在手臂上的场景和痒的感觉联系在一起。大多数这类区域接收的是从多个感官传来的经过了高度处理的输入，而它们的功能原理至今仍是个谜。在本书的后半部分，我还将进一步探讨大脑皮层的层级结构。

在大脑额叶有另一些区域，专门负责运动的输出。大脑皮层的运动系统同样也是层级结构。最低级的运动区M1负责向脊髓传送指令，并直接驱动肌肉的活动。更高级的区域将复杂的运动指令传给M1。运动区域的层级结构和感觉区域的层级结构非常相似，似乎是以同样的方式组织而成。在运动区域，我们认为信息向下传给M1区以带动肌肉；而在感觉区域，我们认为信息在层级结构中以远离感官的方向向上传送。然而实际上，信息同时在向两个方向传递着——感觉区域的反馈正是运动区域的输出，反之亦然。

大多数对于大脑的说明都以流程图作为基础，而流程图所反映的却是一种过于简单的层级结构视图。在这些图中，输入信息（影像、声音和触觉）进入初级感觉区，经过处理后上传至更高层级，随后经过联合区进入大脑额叶皮层，最后被传回至运动区。并不能说这一观点完全错误。当你大声朗读时，视觉信息的确是先进入V1，然后经过联合区，一直传到额叶的运动皮层，最后由运动皮层来控制你的嘴部和喉部肌肉发出说话的声音。然而，这并不是全部的事实，这个过程没有那么简单。我之所以提出警告，是因为这种过于简单的看法认为，信息的传送流程通常被简化为单向的，就像是在工厂的单向装配生产线上制造一个小部件。然而大脑皮层中的信息通常也会反向传输，而且下行的反馈投射比上行投射还要更多。当你大声朗读时，大脑皮层中的高级区域向初级视皮层传输的信号比你的眼睛从页面上接收到的信息还要多！我们将在后面的章节中论及这些反馈投射的作用。现在，我只想让你记住这一事实：尽管信息向上传送的层级结构真实存在，但我们必须提防这样的想法，即以为所有的信息流都是单向的。

让我们再回到解剖台前。假设我们架设了一台强大的显微镜，从大脑皮层上切下了薄薄的一片，给细胞染色后，通过接目镜来观察它。如果将切片中的所有细胞都染色的话，由于大脑皮层中的细胞排列得非常紧密，我们只会看到黑黑的一团。但如果只给一小部分细胞染色的话，我们就会看到前文中所提及的6个层级，每一层在细胞体的密度、细胞类型和连接方式上都存在着不同。

所有神经元都拥有共同的特点。它们除了拥有如你想象的圆形细胞体之外，还有线状的分支结构，称为轴突（axon）和树突（dendrite）。当一个神经元的轴突接触到另一个神经元的树突时，就会形成一些小的连接，称为突触（synapse）。一个细胞的神经冲动就是通过这里去影响到另一个细胞。当一个神经信号的发放到达突触后，很可能会让接受细胞产生动作电位。而有些突触会产生相反的效果，反而让接受细胞产生动作电位的可能性变小。因此，突触可以是抑制性的，也可以是兴奋性的。突触连接的强度可以根据形成它的两个细胞的行为而发生变化。最简单的变化形式是，当两个神经元几乎同时产生动作电位时，它们之间的连接力度就会被加强。这个过程被称为赫布型学习（Hebbian learning），后面我会进一步加以说明。除了改变突触的强度外，还有证据表明，两个神经元之间能够形成全新的突触。这一过程可能随时都在发生，虽然科学上还没有得到一致的证据。无论突触强度发生变化的细节如何，有一点可以肯定，即突触的形成和加强，正是记忆得以存储的原因。

虽然大脑皮层中的神经元分为许多种类，但其中约4/5都属于锥体神经元。顾名思义，它们的胞体形状是锥形的。大脑皮层的6层结构中，除了拥有总长好几千米的轴突和极少数胞体的最顶层结构以外，其余的5层中都包含了锥体细胞。每个锥体神经元都与其邻近的许多其他神经元相连，并将长长的轴突伸向更远处的皮层区域，或下行至丘脑等脑结构。

一个标准的锥体细胞拥有几千个突触。由于它们极端密集且体积很小，我们很难确切知道它们的具体数目，并且，这一数目也会随细胞、层级、区域的不同而变化。如果按照平均每个锥体细胞有1000个突触（实际数量可能接近5000～10000）保守估算的话，那么大脑皮层中就有大约300亿个突触。这样一个天文数字，远远超出了我们能够直观把握的范围。显然，这足够你用来存储毕生所学的东西。

* * *

爱因斯坦曾说，构思狭义相对论并不困难，甚至可以说是轻而易举。它能够从对单一现象的观察中推论得来，即光速对于所有观察者来说是恒定的，即使观察者们在以不同的速度移动。这一点有悖于直觉，就好像说，一个人投球，不管他使用多大的力气，不管投球的速度和对球的观察有多么迅速，被投出的球的运动速度始终是不变的。在任何情况下，任何人所看到的球的速度相对于他们来说都是相同的。这似乎不可能是真实情况。但对光来说，它却被证明是对的。爱因斯坦巧妙地问道，这一离奇的事实会带来怎样的后果？在系统而全面地思考了恒定光速的含义之后，他作了一个关于狭义相对论的更为离奇的预测，即，当你的移动速度加快时，时间就会变慢，而能量和质量从根本上来说是一回事。介绍相对论的书籍使用了诸如火车、子弹、闪光灯等许多日常生活中的例子来帮助说明他的推理过程，可以说，该理论并不难理解，但它无疑是违反直觉的。

在神经科学领域也有一个类似的发现——这个有关大脑皮层的事实如此惊人，以至于一些神经科学家拒绝相信其真实性，而剩下的科学家中大部分人选择对其视而不见，因为他们不知道要如何解读它。然而这一事实如此重要，仔细全面地探索其含义，将为我们提供解开大脑皮层的作用及工作原理之谜的钥匙。这一惊人的事实早就存在于大脑皮层本身的基本解剖结构中，只等着极富洞察力的头脑去发现它。位于巴尔的摩的约翰·霍普金斯大学的神经科学家弗农·蒙卡斯尔（Vernon

Mountcastle）便是拥有这种头脑的人。1978年，他发表了一篇题为《大脑功能的组织原则》（*An Organizing Principle for Cerebral Function*）的论文。在文章中，蒙卡斯尔指出，大脑皮层在外表和结构上存在着惊人的同质性。处理听觉输入的皮层区域与处理触觉的区域相似，同控制肌肉的皮层区域也相似，同布洛卡语言区和几乎所有的皮层区域看起来都一样。蒙卡斯尔写道，既然这些区域看起来都一样，也许他们实际上所起的作用也是相同的！他提出，大脑皮层似乎使用相同的计算方法来完成它的一切功能。

在蒙卡斯尔发表论文的那个年代以及之前的几十年中，解剖学家们全都知道大脑皮层的各个部分看上去相差无几，这是毋庸置疑的。但他们没有去追问这个现象背后的意义，而是花费大把的时间去寻找不同区域之间的差别——也确实找到了一些。他们以为，如果一个区域主管语言，另一个区域主管视觉，那么它们之间就应该存在着差异，只要你够仔细，就一定能发现这些差异。大脑皮层的各个区域在厚度、细胞密度、细胞类型的相对比例、横向连接的长度、突触密度以及许多不易发现的方面均存在不同。初级视觉区V1是被研究最多的区域，在它的其中一层就存在着一些额外的分区。这一情形与19世纪的生物学家所进行的工作有些相似：他们花费了大把时间寻找物种之间的细微差别，最后成功发现，两种外表几乎一模一样的老鼠实际上属于不同物种。达尔文在研究软体动物的那几年里也遵循了同样的方法。但最终他以非凡的洞察力提出了这样的疑问：为何所有物种竟会如此相似？真正令人惊讶和感兴趣的，并非物种之间的差别，而是它们的相似性。

蒙卡斯尔对此也有类似的观察。在解剖学家们皆忙于寻找大脑皮层区域间的细微差别时，他的观察表明，尽管存在着差别，但皮层区域间更多的是惊人的相似。相同的层数、相同细胞类型和连接遍布各处，皮层上的任何区域都像你手中那叠6张的名片一样，其中的差别如此细微，就连训练有素的解剖学家们也难以就其达成一致。因此，蒙卡斯尔认为，大脑皮层所有区域都遵循同样的原理运行，视觉区域之所以

能“看”，运动区域之所以能让肌肉运动，都是由皮层各区域之间以及它们与中央神经系统的其他部分之间的连接方式所决定的。

事实上，蒙卡斯尔认为，造成大脑皮层区域之间细微差别的原因在于它们连接的对象不同，而非它们基本功能的不同。他总结道，所有大脑皮层的区域都有着共同的功能，遵循一个共同的算法，视觉、听觉和运动输出之间并不存在差别。他还认为，大脑皮层区域之间的连接方式是由基因决定的，会因功能和物种的不同而不同，但大脑皮层组织本身在各个区域都做着相同的事情。

让我们再琢磨一下。在我看来，视觉、听觉和触觉是不同的，它们有着迥然相异的基本特性。视觉与颜色、质地、形状、深度和样式有关，听觉则包含音高、节奏和音色。它们所产生的感觉大相径庭，怎么会是相同的呢？蒙卡斯尔说，这些感觉的确是不一样的，但大脑皮层处理从耳朵传来的信号与处理从眼睛传来的信号，所使用的方式是相同的。他接着说，运动控制的工作原理也是同样的。

科学家和工程师们多半忽略了，或者说是无视了蒙卡斯尔的观点。当他们想要理解视觉或是让计算机拥有“看”的能力时，便会去发明一些特定于视觉的专业词汇和技术，将“边缘”、“质地”和“三维表征”等术语挂在嘴边。如果他们想了解口头语言，便会去建立各种基于语法、句法和语义规则的算法。但如果蒙卡斯尔是正确的——这些方法同大脑解决问题的方法完全不同，那么他们将难逃失败的命运。如果蒙卡斯尔是正确的，那么大脑皮层的算法必须独立于任何特定的功能或感觉，即大脑用同样的方式去“看”、去“听”，它的皮层工作原理具有普遍性，适用于任何一种感觉或运动系统。

当我第一次读到蒙卡斯尔的论文时，激动得几乎从椅子上摔下来——仅仅用一篇论文和一个想法就将人类心灵形形色色的奇妙功能联系了起来，这简直是神经科学界的罗塞塔石碑！这些功能被一个简单的算法统一了起来，一击即中地揭示了以往将人类行为分为不同功能加以理

解与设想的尝试其实是一种谬误。我希望大家能够领悟到蒙卡斯尔的观点是多么地超前和优雅。最好的科学观点往往简洁、优雅且出人意料，这一观点就是这样。在我看来，它以前是、现在是，并且很可能将一直会是神经科学领域最重要的发现。然而，令人难以置信的是，大多数科学家和工程师要么拒绝相信，要么选择无视它，甚至有些人根本不知道它的存在。

造成这种忽视的部分原因是研究工具的缺乏。我们没有合适的工具来研究信息流在大脑皮层6层结构中的传递过程。现有的工具都比较粗糙，并且通常只被用于进行各种功能的皮层定位，而不是研究它们的时间特性和模式。例如，当前许多被大众媒体报道的神经科学新闻都隐含着这样的观点，即大脑是一个由高度专业化模块形成的集合体。功能性磁共振成像技术（fMRI）和正电子发射断层成像技术（PET），几乎将重点完全放在前面提到过的大脑分布图和功能区域上。在这些实验中，志愿被试者躺下并将头伸进扫描器内，同时完成一些认知或运动任务。这些任务可能是玩电子游戏、完成动词词形变化、读句子、看许多面孔、命名图片、想象事物、记忆词表、作财务决策等等。扫描器能够检测出在进行这些任务时哪些脑区会比平常更加活跃，并在被试的脑成像图中用不同的颜色将它们标示出来。这些区域因此被推断对完成该任务来说最重要。功能成像实验已经做了上千次，以后还会做上千次。经过这些实验，对于特定功能在成人脑中发生的位置，我们已经逐步建立起了一张地图。你可以轻而易举地说出“这是面孔识别区，这是数学区，这是音乐区”等等。我们并不知道大脑是如何完成这些工作的，因此自然就会假设，大脑在以不同的方式进行各种活动。

但事实果真如此吗？越来越多的证据有力地支持了蒙卡斯尔的理论，一些例子也极好地证明了大脑皮层所具有的非凡灵活性。如果养育良好并置身恰当的环境，任何人的大脑都可以学会世界上几千种口语中的任何一种，同时还能学会符号语言、书面语言、音乐语言、数学语言、计算机语言和肢体语言。它能学会在寒冷的北方气候或在炎热的沙

漠中生活，也可以成为象棋、钓鱼、种庄稼，甚至理论物理方面的专家。请考虑这一事实：在你大脑中有一个特定的视觉区域，似乎是专门用来处理书面字符和数字的。但这是否意味着你的大脑天生就具备了处理文字和数字的语言区呢？恐怕未必。书面语言出现得太晚，基因根本不可能为它专门进化出特定的机制来。因此大脑皮层一直到幼儿时期都还在不断分化为具有特定功能的脑区，而这种分化的基础纯粹是经验。人脑学习和适应上千种环境的能力之强令人难以置信，而这些环境在历史上出现的时间并不长。光凭这一点就足以证明，大脑是一个非常灵活的系统，它不会为一千个问题准备一千种解决方案。

神经科学家还发现，大脑皮层的神经回路具有惊人的可塑性，也即是说，它可以根据输入信息流的类型进行改变和重组。例如，通过手术对新生雪貂的大脑进行重新连线，可以让它的眼睛将信号传送到本该是发展听觉的脑区。这一结果令人惊讶，雪貂在听觉脑区形成了功能完好的视觉通路——它们看东西时所用的脑区在正常情况下本应是用来听声音的。科学家们在其他感官和脑区上也进行了类似的实验。例如，在大鼠刚出生时将它的几块视觉皮层移植到负责触觉的区域，当它发育成熟后，研究者发现，被移植的那些组织处理的是触觉而不是视觉信息。由此可见，神经元在初生伊始并没有专门负责视觉、触觉或听觉的区分。

人类的大脑皮层同样具有可塑性。先天耳聋的成年人，对视觉信息的处理往往发生在原本发展听觉的区域。先天失明的成年人在阅读盲文时，通常使用的是大脑皮层后部的枕叶区，而这个区域通常是主管视觉的。由于盲文涉及手指的触摸，你可能会认为它主要激活的是触觉区，但显然，大脑皮层上的任何区域都不会无所事事。如果视皮层没有接收到“本应”从眼睛传来的信息，它就会从别的皮层区域那里寻找其他的输入模式来加以处理。

所有这些证据都表明，大脑各区域主要根据输入的信息种类来发展出专门的功能。就像地球表面这些国家的区域划分并不是一早注定的一

样，大脑皮层也并不是严格使用不同算法来执行不同功能。你大脑皮层中的组织也如同地球上的政治地理一样，如果在早期设定一个完全不同的环境，就可能会导向与今天完全不同的情况。

基因决定了大脑皮层的整体构造，包括区域之间相互连接的具体细节。但在这个结构的内部，系统却具有高度的灵活性。

蒙卡斯尔是正确的。大脑皮层的各个区域共享着一个强大的通用算法。如果将皮层区域按照合适的层级结构连接起来，并输入一个信息流，它就能学会去了解四周的环境。因此，未来的智能机器并不需要同我们人类一样的感官和能力。大脑皮层的算法能够在人造的机器“皮层”上，通过新的感官和新的方式实现，这样一来，具有灵活性的真正智能便出现了。它将脱离生物脑的范畴。

* * *

下一个与蒙卡斯尔的理论有关的话题，同样会令你惊奇不已，即输入你的大脑皮层的信息基本上都相同。你大概会认为你的感官是完全独立的，毕竟，声音是通过空气传播的压缩波，视觉以光为媒介，而触觉是皮肤上的压力。声音似乎与时间有关，视觉关乎于形象，而触觉主要是空间的。有什么能比羊的咩咩声、苹果的样子和棒球的触感这三者之间的差别更大的吗？

让我们仔细分析一下。从外部世界传来的视觉信息通过视神经中的100万根神经纤维传入大脑。在丘脑经过快速的转换后，信息被传至初级视觉皮层。声音通过听觉神经的3万根神经纤维传入大脑，通过一些较古老的大脑部位后，到达初级听皮层。脊髓通过另外的上百万根神经纤维将有关触觉和内部感觉的信息传入大脑，最后到达初级躯体感觉皮层。这些就是你的大脑的主要输入，也是你对世界的感知。

你可以将这些输入想象为一束电线或者光纤。你应该见过那种由光

纤制成的灯，在每一根光纤的顶端都能看见彩色的光。输入大脑的信息就像是这样，只不过光纤在这里被称为轴突，它们所承载的神经信号被称为“动作电位”，既具有电属性也有化学属性。虽然这些信号来自于不同的感觉器官，然而一旦转化为大脑的动作电位后，它们就成了完全相同的东西——模式。

当一只狗映入你的眼帘时，就会有一组模式通过你的视神经纤维传入到大脑皮层的视觉区。当你听到狗叫声，另一组不同的模式就会通过你的听觉神经进入大脑的听觉区。当你抚摸这只狗时，一组触觉模式就会通过你的手，经过脊椎中的神经纤维，进入大脑中专司触觉的区域。每一个模式——看到狗，听狗叫，抚摸狗——引发的感觉都不同，因为它们在大脑皮层的层级结构中经过的路径不同，而真正重要的正是信息在大脑中的传输路径。但从感官输入的抽象层面上看，它们是基本相同的，都由6层的大脑皮层结构以类似的方式进行处理。你听到声音、看到光亮、感受到皮肤上的压力，但在大脑内部，这些不同类型的信息之间并无根本差异。动作电位就只是动作电位而已，无论源自哪里，这些瞬间发生的动作电位是相同的。你的大脑所能了解的只有模式。

你对世界的所感、所知，都从这些模式而来。你的头脑里没有光，只有一片黑暗；也没有声音，那里一片寂静。事实上，大脑只是你身体的一部分，它本身没有任何感官，即使外科医生将手指伸进了你的大脑，你也不会有所察觉。所有信息都是转化为轴突上的空间－时间模式进入你的头脑的。

空间-时间模式到底是指什么呢？让我们来回顾一下人类的主要感知觉：视觉承载着空间和时间信息。空间模式就是同时发生的多个模式之和，当同一感官的多个感受器受到刺激时，就会形成空间模式。对视觉来说，这一感官便是你的视网膜。一个图像进入你的瞳孔，通过晶状体翻转后投射到视网膜上，形成一种空间模式，再传递给你的大脑。人们通常认为，进入视觉区域的是一个上下翻转的小图像，这其实是一种

误解，并不存在什么图像。所谓的图像，从根本上说，只是一些以特定模式发放的电活动。随着你的大脑对这些信息进行处理，将模式中的各个成分在不同的区域之间传递、筛选和过滤，信息中的图像特性很快就消失殆尽了。

视觉同样也依赖于“时间模式”，也就是说，进入眼睛的模式是随时间不断变化的。虽然视觉的空间模式比较直观易懂，它的时间模式却不是很清晰。你的眼睛每秒钟会快速移动3次，它们起初注视一个点，突然又会跳到另一个点上，这种快速移动被称为扫视。视网膜上的影像会随着你眼睛的每一次移动而不断变化。这意味着，进入你大脑的模式也会随着每一次扫视而彻底改变。这还是当你一动不动地盯着一个不变场景时发生的最简单的情况。在现实生活中，你不断地移动头和身体，行走于不断变化的环境中。你所感觉到的是一个稳定的世界，充满了易于追踪的物体和人。但这种印象完全仰赖于大脑对一系列全无重复的视网膜图像的处理能力。自然的视觉场景所形成的模式就像河流一样进入大脑，与其说它像图画，倒不如说像一首歌。

许多视觉研究者都忽略了眼球扫视和瞬息万变的视觉模式。他们将动物麻醉，观察无意识状态下的动物在注视一个点时如何产生视觉。这样做完全抛开了时间维度的考量。从原则上来说他们做得没错，因为减少可变量是科学实验的核心要素之一。但对于视觉来说，他们所抛开的时间是一个核心成分，是组成视觉的要素。在对视觉的神经科学解读中，时间应该处于中心地位。

提及听觉，我们习惯于考虑声音的时间特性。很明显，声音、口语和音乐都随着时间变化。你不可能即刻听完整首歌，也不可能瞬间听完一个句子。一首歌只在时间的流逝中存在。因此，我们通常不会将声音视为空间模式。在某种程度上，它与视觉的情况正相反：时间模式显而易见，而空间模式则不太明晰。

但听觉同样也有空间的成分。将声音转换成动作电位的是一个叫作

耳蜗的器官，它很微小、不透明、呈螺旋状，镶嵌在人体最硬的骨头——颞骨上。半个世纪前，匈牙利物理学家格奥尔格·冯·贝凯西（Georg von Beksey）解读了它的作用。冯·贝凯西建立了一个内耳模型，他发现，不同的声音频率会引起耳蜗不同部位的振动：高频音引起耳蜗坚硬底部的振动，低频音引起的振动位于耳蜗靠近外部的较柔软的宽阔部，而中频音则振动耳蜗的中间段。耳蜗的每一处都散布着神经元，当被振动时，就会发放动作电位。在日常生活中，你的耳蜗一直都在被大量同时出现的声音频率振动着。因此每一刻耳蜗中都会产生新的空间模式，每一刻都会有一个新的空间模式传到听觉神经。然后，我们再一次看到，这种感觉信息被转化为了空间-时间模式。

人们通常不会将触觉视为时间现象，但它其实既与空间，也与时间有关。你可以做个实验自己判断一下：请你的朋友将手握成杯状，掌心向上，并闭上眼睛。放一个小玩意儿在他的手中——可以是戒指、橡皮擦，什么都行，不允许他移动手的任何部位，让他辨认是什么。这种情况下，除了重量和大致的大小以外，他不会得知任何线索。接下来，仍然让他闭着眼睛，但允许他用手指触摸这个物体，这时他几乎立刻就能识别出来它是什么。出现这一差别的原因，是因为手指的移动为触觉的感知过程赋予了时间。你的指尖就好比视网膜的中心凹一样，两者都极其敏锐。因此，触觉也像是一首歌，你所拥有的与触觉有关的复杂能力，如扣扣子或是在黑暗中开门，全都依赖于这些随时间变化的连续不断的触觉模式。

我们经常教育孩子们说，人有五大感觉：视觉、听觉、触觉、嗅觉和味觉。实际上我们拥有的还要更多。视觉更像是运动、颜色和亮度（黑白对比度）这3种感觉的集合。触觉包括压力、温度、痛觉和振动。我们还拥有一整套的运动传感器系统，即所谓的“本体感受系统”，它能告诉我们关节的角度和身体的位置，没有它，你就无法运动。我们内耳中的前庭系统使我们拥有平衡感。虽然有些感觉相较于其他感觉更为丰富和显著，但它们在进入大脑时，都一样是轴突上随时间变化的空

间模式流。

你的大脑皮层无法直接了解或感知外部世界，它所了解的只有轴突上的输入模式流。你所感知到的世界，包括自我意识，都自这些模式中形成。事实上，大脑无法直接理解生命于何处终结、世界于何处起源这些问题。研究体像（body image）的神经科学家们发现，自我意识比感觉灵活得多。假如给你一个小耙子，用它代替手来抓握，很快你就会感觉到它成为了你身体的一部分。为了适应新的触觉输入模式，大脑会改变它的预期，因此，小耙子已经被合并到你的体像中了。

* * *

来自不同感官的模式在大脑中是相同的，这一观点着实惊人。虽然易于理解，但仍然没有得到广泛的接受。我们可以多举几个实验例子来加以说明。下面的实验你在家中就可以完成，所需要的仅仅是一个朋友、一块遮挡板和一只假手。第一次做这个实验时，最好能找到一只橡胶手，就像万圣节商店里卖的那种，如果没有也没关系，你可以在白纸上描出你的手形。把你的手放在桌面上，与假手对齐，中间相距10厘米（指尖要朝向同一方向，手掌同时朝上或朝下），然后将遮挡板置于这两只手中间的某处，让你的眼睛只能看到那只假手。当你注视着假手时，请你的朋友同时抚摸两只手的相同部位，例如用同样的速度抚摸小指上从指关节到指甲的部位，然后用同样的节奏在食指的第二关节处快速地轻敲三次，接着在两只手背上轻轻画圈，等等。用不了太久，你大脑中视觉和躯体感觉模式相结合的区域（即我前文中所提到的联合区之一）就会变得混乱——你会真实地感觉到从假手传来的触感，就好像它是你的真手一样。

另一个“模式等同”的有趣例子是“感觉替代”（sensation substitution）装置的发明。它可能会彻底改变童年期失明的人的生活，并在不远的将来成为先天性失明患者的福音。除此之外，它还可能为我们这些正常人带来新的机器接口技术。

认识到模式对于大脑的重要性之后，美国威斯康星大学的生物医学工程教授保罗·巴赫-利塔（Paul Bach-y-Rita）发明了一种在人的舌头上显示视觉模式的装置。戴上这一装置的盲人能够通过舌头上的感觉来学会“看”。

它的原理是这样的：在被试的前额上戴一个小型摄像机，并在舌头上放置一块电极板。这一装置能够将所摄图像的像素点一一转换为舌头上的压力点。如此一来，一个可由电视屏幕上的数百个像素构成的粗糙视觉场景，便可以被转换为舌头上由数百个微小的压力点形成的模式并传给大脑，大脑很快就能学会正确地辨认这些模式。

艾瑞克·维汉梅尔（Erik Weihenmayer）是第一批佩戴这种舌上装置的人。他是一位世界级的运动员，13岁时失明。他曾四处演讲，分享与失明命运斗争的经历。在2002年，维汉梅尔登顶了珠穆朗玛峰，成为有史以来尝试并完成了这一挑战的第一位盲人。

2003年，维汉梅尔试用了这个舌上装置，自童年失明后第一次看到了图像。他看到一个球在地板上朝他滚来，还拿起了桌上的一盒软饮，并玩了“石头、剪刀、布”游戏。之后他走向走廊，看到了开着的门，查看了其中一扇门和门框，并注意到门上有一个标志。原本由舌头感觉表征的模式很快转换成了空间图像。

这些实验和例子再一次证明了：大脑皮层是极度灵活的，而进入大脑的输入信息只是一个个的模式而已。这些模式来自哪个感官并不重要，只要它们在时间中以一致的方式相互关联，大脑就能够产生相应的感觉。

* * *

如果我们接受了大脑只了解模式这一观点，所有这一切便都不足为奇了。大脑是处理模式的机器，用听觉或视觉等术语来表述大脑功能并

非不对，但从最根本的层面上来看，模式才是实质。无论各个皮层区域的活动看起来如何不同，它们都基于同一个基本皮层算法。大脑皮层并不关心它所处理的模式是源自视觉、听觉还是其他感官，也不关心它的输入来源是单一感官还是多种感官。无论你是通过声呐、雷达还是磁场来感知世界，或者你身上长的不是手而是触须，甚至你生活在四维而非三维世界里——这些大脑皮层统统不在意。

这意味着，要想具有智能，你不必需要任何一个感官或感官间的特定组合。海伦·凯勒（Helen Keller）既看不见也听不见，然而她学会了语言并成为了一位比大多数耳聪目明的人更为出色的作家。作为一个缺失了两种主要感觉的智能人类，她的大脑以惊人的灵活性使她能够像感官健全的人一样去感知和认识世界。

人类头脑的这种非凡的灵活性，让我对基于大脑研究的未来技术满怀希望。当我考虑建造智能机器时，我就会想，为什么要固执于我们所熟悉的感觉上呢？只要能破译新皮层的算法并创立一种模式科学，我们就可以将其应用到任何想使之拥有智能的系统上。这种由大脑皮层所启发的回路，其重要特点之一就是，我们不需要特别聪明才能操控它。在被动了手脚的雪貂大脑中，听觉皮层可以变为"视觉"皮层；盲人大脑中视觉皮层的功能可以被替代；同样地，一个以大脑皮层算法运行的系统，将在我们选择输入的任何模式的基础上表现出智能。当然，我们仍需要聪明才智来设置广泛的系统参数，仍然需要训练并调教它。但是之后，令大脑拥有复杂和创造性思维能力的那数十亿神经单元将会自然形成，就像在儿童身上所发生的那样。

最后，模式是智能的基本媒介这一观点，带来了一些有趣的哲学问题。当我和朋友们坐在一个房间里时，我怎么知道他们在那里，甚至怎么确定他们是不是真实存在的呢？我的大脑接收到一组同过去经历过的模式相一致的模式。这些模式对应着我所认识的人——他们的脸、他们的声音、他们的言谈举止以及各种方面。我已经学会并习惯了对这些以

可预见的方式同时发生的模式产生期待。但是，说到底，这一切都只是个模型。我们对这个世界的所有认识都是建立于模式之上的模型。我们能肯定这个世界一定是真实的吗？这是个奇怪而有趣的问题，科幻小说和电影中时常探讨它。并不是说人或物体不是真的存在——他们确实存在，而是说，我们对于世界存在的肯定，是基于模式的一致性以及我们对它们的解读上的。直接的感知根本不存在，我们没有相应的感受器可以用来感知"人"。请谨记：大脑是一个寂静的黑盒子，内部除了输入纤维上随时间流动的模式之外，别无他物。你对世界的认识是由这些模式而不是别的什么东西创造的。存在可能是客观的，而那些不断流向大脑中轴突束的空间——时间模式，才是我们人类所必然经历的。

这些讨论令我们注意到一个时常困扰我们的问题，即幻觉和现实的关系问题。如果你可以幻想出橡胶假手上的感觉，并可以通过舌头上的触觉刺激"看"到图像，那么你在自己的手上感觉到的触摸和亲眼所见的一切，会不会也是由类似的"欺骗"所产生的幻觉呢？我们能相信世界就是我们眼前的样子吗？我相信是。世界确实以一种绝对形式存在着，并且和我们所感知到的非常接近。然而我们的大脑无法直接认识到那个绝对的世界。

大脑认识世界所凭借的一系列感觉，只能探测到绝对世界的一部分。这些感觉形成的模式被传入大脑皮层，由同样的皮层算法处理后，创建出世界的模型。如此一来，虽然口语和书面语在感觉层面上完全不同，但被感知的过程却惊人得相似。因此，尽管海伦凯勒有着严重的感觉缺失，但她头脑中的世界模型与你我相差无几。大脑皮层通过这些模式构建了一个近乎真实世界的模型，然后将它存放于记忆之中。在下一章中，我们就将来探讨记忆，看看模式在进入皮层之后发生了什么。

第四章 记忆

无论你是正在读这本书，还是行走于拥挤的街道；无论你是在听交响乐，还是在安慰哭泣的孩童，你的大脑中都充满了从身体各个感官传来的空间与时间模式。

世界就像是一个不停变化的模式的海洋，不停地拍打和冲刷着你的大脑。你是如何理解这些冲击的呢？模式不断涌入，经过旧脑的各个部分，最后到达新大脑皮层。然而在它们进入大脑皮层之后，又发生了什么？

自工业革命初，人们就将大脑视为某种机器。虽然他们知道头脑中并没有齿轮和螺丝钉，但这是当时所能想到的最好比喻——信息以某种方式进入大脑，然后由大脑这个机器来决定身体应当作何反应。在之后的计算机时代中，大脑又被看作另一种特别的机器——可编程计算机。正如我们在第一章中所提到的，人工智能的研究者们始终坚持这个观点，并认为他们的研究之所以缺乏进展，只是由于计算机与大脑相比体积太小、速度太慢。他们说，现今的计算机只相当于一只蟑螂的大脑。如果能造出更大、更快的计算机，它们就会像人一样聪明。

这个大脑与计算机之间的类比忽略了一个极为重要的问题。与计算机中的晶体管相比，神经元要慢得多。神经元从突触中收集输入信息，并将它们结合起来，以决定何时向其他神经元输出电脉冲。一个典型的神经元可以在5毫秒内做到这一点并自行复位，大约相当于每秒200次。这看起来似乎很快，但一台现代硅芯片计算机可以在1秒内完成10万次运算。这意味着，计算机的一次基本运算的速度，要比你的大脑快500万倍。这是一个非常非常大的差异。大脑怎么可能比最快的数字电脑更

快、更强大呢？坚信大脑类似于计算机的人会说，“大脑是一个并行计算机，它有几十亿的神经元同时进行运算。这种并行性大大增强了生物大脑的处理能力。”

我一直觉得这是一个谬论。只需一个简单的思维实验就能说明我的观点，这个实验叫作“一百步法则”。一个人可以在不到一秒的时间内完成大量任务。例如，我给你看一张图片，并让你判断其中是否有一只猫，如果有就按下一个按钮，如果看到的是一只熊、疣猪或者萝卜，就不要按。对于今天的计算机来说，这是个困难的任务，甚至不可能完成；而一个人却可以在半秒甚至更短的时间内轻松做到。但是神经元传导是缓慢的，所以在半秒内，进入你大脑的信息只能传过100个神经元长度的链条。也就是说，不管总共有多少神经元参与，大脑总能在“一百步”之内“计算”出类似问题的解决方案。从光线进入你的眼睛到你按下按钮的这段时间内，参与整个过程的神经元不会超过100个。试图解决同样问题的数字计算机，则需要走几十亿个步骤。100个计算机指令仅够在计算机屏幕上移动单个字符，更别提做什么有趣的事情了。

但是，如果几百万的神经元共同工作，会不会类似于一个并行计算机呢？也不尽然。大脑以并行方式运行，并行计算机也是，这是它们唯一的共同点。并行计算机将多台快速计算机结合在一起来处理复杂的问题，比如计算明天的天气情况。要预测天气，你必须计算地球上许多地理位置上的物理状况。在相同的时间点上，每台计算机都负责一个不同的地理位置。然而，即使有数百甚至上千台计算机并行工作，每一台计算机仍然需要运行几十亿甚至几万亿个步骤才能完成它们的任务。人类所能想到的最大的并行计算机，也无法在“100步内”做出任何有用的事情，无论它有多大、速度有多快。

这儿有一个类比。假设我让你将100块大石头从沙漠的一边运送至另一边。你一次只能携带一块石头，而穿越沙漠需要走一万步。你琢磨着，如果只凭自己的话将花费很长的时间，于是你雇了100名工人来一

起运送。现在工作速度提高了100倍，然而穿越沙漠仍然需要至少一万步。雇用更多的工人，甚至上千名，也不会再有任何的提高。无论你雇用多少工人，一个需要一万步的工作也无法在更短的时间内完成。这个道理对于并行计算机来说也是一样。超过某个数量之后，增加再多的处理器也不会有所增益。一台计算机，无论它有多少个处理器，无论它运行的速度有多快，也不可能在“一百步”内“计算”出复杂问题的答案。

那么，大脑是如何能在“一百步”内处理那些即使由最大的并行计算机用100万甚至几十亿个步骤也解决不了的复杂任务的呢？答案很简单：大脑并不“计算”问题的答案，它是从记忆中提取答案。这些答案实际上是很久以前被存储在记忆中的，只需要几个步骤，就可以从记忆中提取出来。缓慢的神经元们，不仅足以胜任这一工作，而且本身就构成记忆。整个大脑皮层就是一个记忆系统，根本不是什么计算机。

* * *

让我用下面这个例子来说明，“用计算”和“用记忆”来解决同一问题，这两者之间有何不同。现在你的任务是接住一个球。有人向你投来一个球，你看着它飞向你，在不到1秒的时间内，你在空中将它接住。这看起来似乎并不太难——而如果你尝试编写一个机器人手臂的程序让它做同样的任务，就会发现情况比你想象的要复杂得多。正如许多研究生们以惨痛的经历所了解到的那样，这几乎是不可能完成的任务。工程师或计算机科学家们在解决这个问题时，他们首先会尝试计算球飞行的路径，以确定当它到达手臂时的位置。这个计算需要求解一系列你在高中物理课上学过的那种方程。接下来，机器人手臂上的所有关节都要协调一致地运作，以使手移动到正确的位置。这一步骤又涉及比之前更复杂难解的另一组数学方程。最后，这整个过程必须重复多次，因为随着球越飞越近，机器人会更加精确地获知它所处的位置和飞行轨迹。而如果等到球飞到准确的位置上才开始有所行动的话，就太晚了。机器人必须在还不十分清楚球的位置时就开始移动，并随着球越来越近，不断地

进行调整。要接住这只球，一台计算机需要运行几百万个步骤，以求解大量的数学方程。而即使计算机程序有可能成功地解决这一问题，“一百步法则”也让我们明白，大脑在解决这个问题时使用的是不同的方式，那就是记忆。

你如何使用记忆接住球呢？你的大脑内存储着对接球所需的肌肉指令的记忆（还有许多其他的习得行为）。当球被抛出后，会有3件事发生：首先，球的影像会唤起相关的适当记忆。其次，这个记忆实际上会引发出一个肌肉指令的时间序列。最后，被提取的记忆会依据当时的特定情况，如球的实际轨道和你的身体位置，来进行调整。如何接球的记忆并不是被编入大脑的程序，而是通过多年的反复练习学习到的，它存储于你的神经元中，而不是基于神经元的计算。

你可能会想：“且慢。每次接球的情况都会略有不同，你刚才说，被提取的记忆会不断调整，以适应任何一次投球时球的位置变化……那岂不是仍然需要面对我们曾试图避免的方程吗？”看起来的确如此，但大自然以一种非凡的方法，解决了这一问题。我们将在本章的后面看到，大脑皮层创建了一种叫作恒定表征的能力，能够自动处理这种变化。想象一下当你坐在水床上的情形：水床上的枕头和人都会不由自主地被推到一个新的位置。床本身并没有计算每个物体应该被抬高多少，水的物理性质和床垫的弹性表面会自动调节这一切。在下一章中我们将看到，6层结构的大脑皮层，不严格地说，对于流经它的信息具有类似的作用。

* * *

因此，大脑皮层并不像计算机，无论是并行的还是其他类型。它不会去计算问题的答案，而是用存储的记忆来解决问题并产生行为。计算机也有记忆，以硬盘驱动器和内存芯片的形式；但大脑皮层的记忆有4个属性是完全区别于计算机记忆的：

- 大脑皮层存储的是序列模式。
- 大脑皮层以自-联想的方式提取模式记忆。
- 大脑皮层以恒定的形式存储模式。
- 大脑皮层将模式存储在层级结构中。

我们将在本章讨论前3个区别。在第三章中，我介绍了大脑皮层的层级结构。在第六章，我将继续论述它的重要性以及工作原理。

下一次当你讲故事时，不要着急开始，先退一步想想，如何才能一次只讲述故事的一个方面。因为不论你的语速有多快，也不论我的理解有多快，你都无法将发生的所有事情一次性告诉我。你需要先讲一部分，再接着讲下一部分。这样做不仅因为口语是连续的——不论书面、口头，还是视觉叙事都是以序列的方式来传达故事；还因为故事是以序列的方式存储在你的大脑中，并且只能以相同的顺序被回忆起来。你无法一下子记起整个故事。事实上，你几乎不可能想起任何非序列化的复杂事件或想法。

你或许已经注意到，有些人在讲故事时不能立刻切入重点，而是东拉西扯地讲一些无关紧要的细节。这会让人很恼火，想要大叫说："快讲重点吧！"但他们之所以如此，是因为这个故事在时间上就是这样发生在他们身上的，他们无法以其他方式来讲述它。

另外一个例子：请闭上眼睛想象一下你的家。你的想象中，走到前门。想象一下它的样子。打开前门，走进去。现在看看你的左边，你看到了什么？再向右看，又看到了什么？走到你的浴室。右边是什么？左边是什么？右上抽屉里有什么？在你的浴室里放了些什么物品？你知道这一切，甚至更多，并能清楚地回忆起它们。这些记忆都储存在你的大脑皮层中。你可能会说，这些东西都是关于你家的记忆的一部分。但你不可能一次将它们全部记起来。它们显然是相关的记忆，但你做不到将

全部细节一下子全记起来。你对家有一个完整的记忆，但要想起它，你必须按照连续的场景回忆，就像你经历它时那样。

所有的记忆都是这样。你必须经历事件发生时的时间序列。一个模式（走近门）唤起下一个模式（进入门），这个模式又唤起下一个模式（穿过大厅或拾级而上）等等，依此类推。每一个记忆都是你曾经历过的序列。当然，通过有意识的努力，我可以改变描述我家的顺序。如果我决定以非顺序的方式重点介绍一些项目，就可以从地下室直接跳至二楼。然而，一旦我开始描述我选择的任意房间或项目时，我又会按照一定的顺序来讲。真正随机的想法是不存在的。回忆总是沿着一条联想的路径展开。

你一定很熟悉字母表，那么请试着倒背它。你会发现你做不来，因为你并不经常从后往前地读它。如果你想知道小孩子学习字母表是什么感觉，试着把它倒背一下，那就是他们的感觉。的确很难。你对字母表的记忆是一个由模式构成的序列，而不是那种能在瞬间以任意顺序存储或调用的东西，就像一周中的每一天，一年中的每个月，你的电话号码，以及无数其他事物一样。

你对歌曲的记忆是用来说明记忆中时间序列的极好例子。回想一首你知道的曲子，我喜欢用《彩虹之上》（*Somewhere over the Rainbow*），其他曲子当然也可以。你会发现，你无法一时间想起整首歌曲，而只有按照顺序依次回忆。你可以从头或者副歌开始，一个音符接一个音符地唱下去。就像你不能一下子记起全部一样，你也不能从后往前唱。在你第一次听到《彩虹之上》的时候，它是在时间中依次播放的，因此你只能按照你学歌时的顺序记起它。

这同样适用于低级感觉记忆。比如你对质地的触觉记忆。你的大脑皮层记得手中满握一把沙砾的感觉，手指滑过天鹅绒的感觉和按下钢琴琴键的感觉。这些记忆同字母表和歌曲完全一样，以序列作为基础，只不过这些序列更短，只持续几分之一秒，而不是几秒或几分钟。如果在

你熟睡时将你的手埋在一桶砾石里，当你醒来时，就不会知道自己摸到的是什么，除非你移动手指。你对于砾石的质感触觉记忆，是建立在由皮肤上的压力和震动感受神经元形成的模式序列之上的。这些序列完全不同于将手埋在沙子、泡沫小球或干树叶中所感受到的模式序列。当你弯曲手指的时候，小石头的摩擦和滚动会形成砾石特有的模式序列，从而触发在你躯体感觉皮层中的相应记忆。

下一次洗澡时，请留意自己是如何用毛巾擦干身体的。我发现自己几乎每次都是以完全相同的顺序，并配以相应的姿势来擦、拍身体。通过一个有趣的实验，我发现我的妻子在淋浴之后也遵循着一个半固定的模式。你很可能也是这样。如果你遵循着一个顺序的话，试着改变它。你会发现，虽然可以要求自己改变，但你需要时刻保持专注。因为如果稍一分神，就会回到你习惯的模式上去。

所有的记忆都被存储在神经元之间的突触连接中。我们在大脑皮层中存储了非常多的信息，而在任一时刻我们可以回想的只占其中的一小部分，由此可见，在每一次记忆提取过程中，只有有限的突触和神经元发挥了积极的作用。当你开始回想家里都有什么时，一组神经元开始变得活跃，接着另一组神经元会被它们激活，依此类推。一个成年人的大脑皮层，拥有大到惊人的记忆容量。但尽管我们已经存储了非常多的信息，也只能一次记起其中的一小部分，而且只能按照联想顺序来回忆。

这里有一个有趣的练习。努力回想一下有关自己过去的细节——你曾经的住处、曾经去过的地方、曾经认识的人。我发现自己总能揭开那些尘封多年未曾触碰过的记忆。在我们大脑的突触连接中，存储着成千上万个几乎从未用过的记忆细节。无论何时，我们的回忆只是其中一个很小的片段，而大部分的信息则存在那里，静静地等待着合适的线索来调用它们。

计算机的内存通常不会存储模式序列。你可以使用各种软件来实现这种存储（比如在计算机里存一首歌），但计算机内存本身并不会自动

完成这一工作。与此相反，大脑皮层则会自动存储序列，这是大脑皮层记忆系统的一个固有特性。

* * *

现在，让我们来看看记忆的第二个主要特征，即它的自-联想特性。正如我们在第二章中看到过的，这个术语即是指模式与自己相关联。自-联想记忆系统，就是一个能够根据不完整或被扭曲的输入信息，提取出全部完整模式的系统。它对于空间和时间模式都适用。如果看到你的孩子从布帘后面露出来的鞋，你就会自动联想起他或她的整个样子。你从鞋子的这一部分影像出发，补全了由孩子的整个影像所形成的空间模式。或者想象你看到的一个正在等公交车的人，因为她站在灌木丛后面，你只能看到她的一部分，但你的大脑并不会因此混乱。虽然眼睛只看到她身体的几个部位，但你的大脑会自动填补余下的部位，创造出一个完整的人的知觉。这种知觉是如此强烈，你甚至意识不到它只是你的臆断。

你也可以填补时间模式。如果回想在很久以前发生的某件事的一个小细节，与此相关的整个记忆序列就会像潮水一样涌入你的大脑。马塞尔普鲁斯特（Marcel Proust）的著名系列小说《追忆似水年华》，开篇就是对玛德琳点心香味的回忆，之后便一发不可收拾地写了1000多页。在嘈杂的环境中，我们经常听不清所有的谈话内容，但这并不妨碍，我们的大脑会用它所期望听到的来填补那些没有听到的内容。可以确定的是，我们并没有真实听到所有被知觉的话语。你一定见识过那些喜欢大声接别人话的人，而其实在我们的大脑深处，每一个人都在不停地这样做，而且不只是接句子的末尾，连中间和开头也接。在大多数时候，我们都意识不到自己正在不断填补模式，但它确实是皮层存储记忆的一个无处不在的基本特征。无论何时，一个记忆片段都可以激活全部的记忆，这便是自-联想记忆的本质。

你的大脑皮层是一个复杂的生物自-联想记忆系统。在你清醒的每

个时刻，所有功能区都在警觉地等待熟悉的模式或模式片段的输入。在你沉思的时候，一个朋友的出现会让你的思绪瞬间切换到她的身上。你并没有主动选择这一切换，但你朋友的出现会迫使你的大脑开始回忆与她相关的模式。这一反应是无法遏制的。被这个干扰中断了思路的我们经常会问自己："我刚才在想什么来着？"与朋友的晚餐交谈也是一条迂回联想的路线，谈话可能会从你面前的食物开始，然后桌上的色拉会让你想起妈妈在你婚礼上做的色拉，而这又会让你们想起另一个人的婚礼，并由此联想到他们度蜜月的地方，以及那个地方的政治问题，等等。思想和记忆是相互关联的，我要再一次强调，随机的想法是不可能发生的。输入到大脑的信息会自动关联自己，填充当前信息，并与通常接下来要发生的事情联系起来。我们将这一连串的回忆称为思想，虽然我们其实并不能确定它的路径，亦无法完全控制它。

* * *

现在让我们来看看大脑皮层的第三个主要特性：它是如何形成所谓的恒定表征的。我将在本章中介绍有关它的基本概念，在第六章中，我还会详细解释大脑皮层如何创造出它们。

计算机内存中存储的信息就是信息本身所展现的那样。如果你从光盘上拷贝一个程序到硬盘，那么它的每一个字节都将和之前的百分之百相同。两个版本之间哪怕有一丁点的错误或差异，都可能会导致程序崩溃。而大脑皮层中的记忆则与此不同。我们的大脑并不确切记得它的所见、所听和所感。我们之所以不能完全精确地进行记忆或回忆，并不是因为大脑皮层和它的神经元做事马虎或是易出纰漏，而是因为大脑所记忆的是这个世界上各种独立于细节的重要联系。让我来举几个例子说明这一点。

在第二章中我们曾提到，简单的自-联想记忆模型已经存在了几十年，而我在本章前面也说过，大脑以自-联想的方式提取记忆。然而，神经网络研究者们所建立的自-联想记忆同大脑皮层的记忆之间有着很

大的区别。人造的自-联想记忆没有使用恒定表征，因此它们在一些极为基础的方面表现得很失败。设想我有一张由大量黑白圆点组成的人脸图片，这幅图即是一个模式，如果我有人工的自-联想记忆，就可以在记忆中存储许多这种类型的脸。如果我呈现给它半张脸或者只是一双眼睛，它就会识别出这是图像的一部分，并将缺少的部分正确地填补上去。这正是它的擅长之处。这种类型的实验已经做过许多次。然而，如果我将图中的每个圆点向左移动5个像素，自-联想记忆系统就完全认不出这张脸了。对于人造的自-联想记忆来说，它变成了一个全新的图片，因为原来存储的模式同新模式的像素无法吻合。然而你我却可以毫不费力地将偏移后的模式视为同一张脸，甚至根本不会注意到其中的变化。如果模式被移动、旋转、重新调整了大小或是产生任何其他改变，人工智能的自-联想记忆就会无法辨认出它们；而我们的大脑却对这些变化应付自如。当表征一件事物的输入模式产生变化时，我们怎么就能将它感知为是相同或恒定的呢？让我们来看另外一个例子。

此刻你手里应该正拿着一本书。当你移动这本书，或是改变照明、调整坐姿、注视页面上的不同部分时，投射在你视网膜上的光的模式就会完全改变。你接收到的视觉输入每时每刻都在变化，并且从不重复。事实上，即使你将这本书捧上100年，投射到你视网膜上的模式以及之后进入你大脑的模式中也不会有任何两次是完全相同的。然而，你丝毫不会怀疑你正捧着一本书，而且还是同一本。尽管接收到的刺激不断变化，大脑中表征“这本书”的内部模式却不会改变。正因如此，我们才使用“恒定表征”这一术语来指代大脑的内部表征。

再举一例。请回想一位朋友的脸。每次当你见到这张脸，就能认出她来，这一过程在1秒钟内自动发生，无论她距离你是有一两米，还是在房间的另一端。当她离你很近时，她的影像会占据你大部分的视网膜；当她离你很远时，她的影像只会占据你视网膜的一小部分。她可以正面朝向你，也可以稍微侧一点，或是侧面朝向你；她可以面带微笑，微微眯眼，或打着哈欠；她可以在明亮的光线中，也可以在阴影里，或

在迪斯科舞厅里角度奇怪的灯光下；总之，她的面孔可以出现在无数位置上，伴以无数的变化。尽管投射在你视网膜上的光线模式每次都不尽相同，但每次你都能立刻知道自己注视的是她的脸。

让我们揭开帘幕，看看在执行这一惊人的功能时，你的大脑里究竟发生着什么。如果监测大脑皮层视觉输入区（即V1区）的神经元活动，实验会显示，她脸的每个不同影像所引发的活动模式都不一样。随着脸的每一次移动，或是你眼睛的每一次新的注视，V1中的活动模式都会发生变化，与视网膜上的模式变化类似。然而，当监测你的面孔识别区时，我们却发现，这一高于V1好几层级的功能区中的细胞活动非常稳定。也就是说，只要你朋友的脸出现在你视野中的某处（甚至是在想象中），不论其大小、位置、朝向、比例和表情如何，面孔识别区的一些细胞就都会保持活跃。细胞兴奋发放的这种稳定性就是一种恒定表征。

从内省的角度来看，这个任务似乎简单到不值一提，就像呼吸一样自然。而它之所以看起来微不足道，原因之一是我们察觉不到它的发生。或者在某种意义上说，它之所以微不足道，是因为我们的大脑可以很快地解决它（别忘了“一百步法则”）。然而，“大脑皮层如何形成恒定表征”这一问题，仍然是所有科学问题中最大的谜团之一。你可能会问，这能有多难呢？——难到即使动用世界上最强大的计算机，也没有人能够解决，并且尝试解决它的人也非常之多。

对这个问题的思索由来已久，可以一直追溯到2300年前的柏拉图。他曾困惑于人们如何能够思考和了解世界。他指出，真实世界中事物的和概念的实例总是不完美和不相同的。比如，你对完美的正圆有一个概念，但你从来没有真正见过正圆，因为所有画出的圆都是不完美的。即使用几何学家的圆规画出的所谓的圆，也有一条黑线的边，而一个真正的圆的周边是没有厚度的。你是从哪里获得正圆这个概念的呢？或许我们可以举另外一个更为生活化的例子，想想你对狗的概念。你见过的所

有狗都不尽相同，即便是对同一条狗，每次也都会看到不同的影像。所有的狗都不相同，你也不可能以完全相同的方式再次看到同一条狗。但是你对狗的所有不同经验会汇集为一种叫作“狗”的稳定的心理概念。柏拉图对此困惑不已。在这个拥有无限种形式和千变万化的感觉的世界上，我们是怎么学会并运用概念的呢？

柏拉图提出的解决方案，即是著名的理念论。他总结道：我们的高级心智一定是被束缚在超现实的某些先验层面上，其中存在着永恒完美的稳定概念（即理念，以大写字母F代替）。他认为，我们出生前的灵魂就来自于这个神秘之处，并在这里第一次获得各种理念。在我们出生后，仍然保留着这些潜在的知识。学习和理解的发生，是因为现实世界的形式使我们回忆起了对应的理念。你之所以知道“圆”和“狗”的概念，是因为它们分别触发了你灵魂深处对“圆”和“狗”这些理念的记忆。

从现代的角度来看，这个理论显然很疯狂。然而，如果去除掉那些夸张的形而上学成分，你就会发现，他实际上是在谈论恒定性。尽管他的解释系统完全不着边际，但他的直觉却道出了问题的关键：恒定性正是有关人类本质的最重要问题之一。

* * *

为了避免让你产生恒定性只与视觉有关的错误印象，让我们再来看一些其他感官的例子。首先来看你的触觉。当你伸手到汽车的手套箱中寻找太阳镜时，你的手指只需稍一触碰，就知道是不是找到了。无论接触部位是你的拇指，还是指尖的任何部分或手掌；也无论接触到的是镜片、镜腿、铰链处，或是镜框的一部分；你手的任何部位在太阳镜的任何部分上只需滑过1秒钟的时间，就足以让你的大脑识别出它是太阳镜。在每种情况下，来自触觉感受器的空间与时间模式流都完全不同——不同的皮肤区域，不同的物体部位——然后你会不假思索地将太阳镜一把抓起。

再来看一个感觉运动的任务——将钥匙插入汽车的点火开关。每次做这个动作时，你的座位、身体、手臂以及手的位置都会略有不同。但对你来说，它就是一个日复一日的简单重复的动作，因为在你的大脑中存储着这一动作的恒定表征。如果你试图造一个能走进车里并插入钥匙的机器人，很快你就会发觉这几乎是不可能的，除非你能确保机器人每一次都处于完全相同的位置，并以完全相同的方式拿着钥匙。即使你设法做到了这一点，也还要根据不同的汽车给机器人设计不同的程序。机器人和计算机程序，如同人工自-联想记忆一样，在处理变化方面都十分蹩脚。

另一个有趣的例子是你的签名。在你大脑额叶的运动皮层某处，有一个亲笔签名的恒定表征。每次签名时，你使用的笔画、角度和节奏的序列都是相同的。无论是用是细头钢笔仔细描画，还是像约翰·汉考克（John Hancock）那样用胳膊肘在空中华丽地一甩，或是用脚趾夹着铅笔笨拙地签名，情况是完全一样的。当然，你的签名每次看起来都有些不同，尤其是在我刚才所说的一些尴尬情形下。然而，无论字的大小如何，书写工具以及所使用的身体部位怎样，你总是以相同的抽象“运动程序”来完成签名这一动作。

从签名的例子中可以看出，运动皮层中的恒定表征，在某些方面与感觉皮层中的恒定表征互为镜像。在感觉方面，多种输入模式可以激活表征某些抽象模式（朋友的脸、太阳镜）的一组稳定细胞群，在运动方面，表征某些抽象运动指令（接球、签名）的一组稳定细胞群又能够使用各种各样的肌肉群并遵照各种各样的约束来表现自己。如果蒙卡斯尔的观点是正确的——大脑皮层在各个区域使用的是同一个基本算法，那么感觉与运动之间的对称性就正是我们所期望看到的。

最后，让我们回到感觉皮层，再看看与音乐有关的例子（我喜欢以音乐记忆为例，因为它能令我们轻易看到大脑皮层必须解决的一切问题）。音乐中的恒定表征体现为，你能认出以任何调式演奏的同一段旋

律。乐曲的演奏调式是旋律的基本音阶。以不同调式演奏的旋律起始于不同的音符。一旦选择了演奏的调式，也就确定了旋律中的其他音符。任何一首旋律都可以用各种调式演奏。这意味着，用新的调式演奏的同一首曲子，实际上是完全不同的一个音符序列！每次演奏都在耳蜗的一系列完全不同的位置上产生刺激，由此引发一个完全不同的空间-时间模式流传入你的听觉皮层，然而每次你感受到的都是同样的旋律。除非有完美的音准能力，否则，如果不连续听，你甚至无法区分曲子是由两种调式分别演奏的。

想想那首《彩虹之上》吧。你第一次听到它，可能是在电影《绿野仙踪》（*The Wizard of Oz*）中由朱迪·加兰演唱。然而除非有完美的音准能力，否则你不可能记得她所唱的调式（降A调）。如果我坐在钢琴旁，以你从未听过的调式（比如D调）来演奏这首曲子，听起来会是同一首。你不会注意到所有的音符与你熟悉的那个版本已经完全不同了。这意味着，你对这首歌曲的记忆一定是一种与调式无关的形式。记忆存储的必然是歌曲中的重要关系，而非实际音符。这里所说的重要关系，是指音符的相对音高，或称“音程”。《彩虹之上》以一个高八度开始，接着降了半调，然后是一个降大三度……同一旋律的音程结构对于以任何调式演奏的版本都是相同的。能够轻易识别出任何调式的同一歌曲的能力表明，在你的大脑中，歌曲是以恒定音高的形式存储的。

同样，对你朋友的脸的记忆，也必然是以一种独立于任何特定影像的形式来存储的。你凭借脸的相对大小、颜色和比例而认出她，而不是她在上周二午餐时某一个瞬间样子。就像一首歌的音符之间有“音高间隔”一样，她脸的特征之间也存在着“空间间隔”。她的脸相对她的眼睛而言比较宽。她的鼻子相对于眼睛的宽度来说比较短。她头发的颜色和眼睛的颜色也有一种类似的相对关系，即使在不同的光照条件下，在绝对颜色改变明显的情况下，这种关系也是保持不变的。当你记忆她的脸时，所记的就是这些相对特征。

我相信，在大脑皮层的每个区域上都发生着类似的形式上的抽象。这是大脑皮层的普遍特性。记忆存储的是关系的本质，而不是片刻的细节。当你在看、在感觉或是在听的时候，大脑皮层接收了高度特异的详细输入信息，并将其转换为一种恒定形式。被存储在记忆中的是恒定形式，与每一个新的输入模式相比较的也是恒定形式。记忆的存储、提取和识别都发生在恒定形式之上。而计算机中没有与之等效的概念。

* * *

这就引出了一个有趣的问题，在接下来的一章，我提出了这样一个假设：大脑皮层的一个重要功能是利用记忆进行预测。但既然大脑皮层存储的是恒定的形式，它要如何作出具体的预测呢？下面的一些例子可以用来说明这个问题及其解决方案。

想象现在是1890年，你居住在美国西部的一个边陲小镇。你的心上人正从东部坐火车赶来与你开始共同生活。你当然想在她到达时去车站接她，于是你在她来之前的几个礼拜，就提前关注了火车时刻表。但当时没有固定的时刻表，你所能了解到的只是，火车在一天之中不会在同一时刻到达或离开。你开始觉得似乎无法预测她的火车将于何时到达。但随后你注意到，火车的到站和离站时间中有一些规律。自东部来的火车的到站时间，要比往东去的火车离站时间晚4个小时。虽然每天到站和离站的具体时间点变化很大，但中间这4个小时的间隔是日日相同的。在她到达的那一天，你只需留意向东去的火车，当看到它离站时，便设定你的时钟。4个小时后，你起身前往车站，那时她乘坐的火车刚好到达。这个例子说明了大脑皮层所面临的问题，同时也告诉了我们大脑是如何解决的。

你的感官所感受到的世界永远不会相同，就像火车到达和离开的时间一样。你只有通过在不断变化的输入流中寻找恒定的结构来了解世界。然而，只拥有这种恒定的结构，也不足以让你做出具体的预测。这就像是，只知道列车在上一趟列车离站后4小时到达这一信息，并不能

让你准点出现在站台上迎接你的心上人。为了做出准确的预测，大脑必须将对于恒定结构的了解同最新的细节信息结合在一起。预测列车的到达时间，需要了解列车时刻表的4小时间隔，并将它与细节信息——最近一趟东去列车的离站时间相结合。在聆听一首熟悉的钢琴曲时，你的大脑皮层在下一个音符演奏之前就已经给出了预测。但对于曲子的记忆，如前所述，是以音高恒定的形式存储的。你的记忆能告诉你下一个音程是什么，但却不能告诉你下一个实际的音符是什么。想要准确地预测下一个音符，需要将下个一音程同你所听到的最后一个音符结合起来。如果下一个音程是大三度，而所听到的最后一个音符是C调，你便可以由此预测出下一个音符是E调。你的脑中听到的是E，而非大三度。除非你认错了曲子，或者钢琴师弹错了，否则你的预测一定是正确的。

当你看到朋友的脸时，你的大脑皮层会在一瞬间补充并预测关于她独特形象的万千细节，检查她的眼睛是不是无误，鼻子、嘴唇和头发也都是应该有的样子。你的大脑皮层所作出的预测具有极大的特异性，它可以预测出她的脸的底层细节，甚至在你从来没有看过的角度和环境下的细节。如果你对你朋友眼睛和鼻子的位置非常熟悉，并且了解她脸的结构，那么你就可以准确预测出她嘴唇的位置。如果你知道她的皮肤被晒成了橘色，那么你就可以推测出她的头发看起来应该是什么颜色。你的大脑正是将对脸的恒定结构记忆同你即时经验的详情相结合来做到这一点的。

列车时刻的例子只是对大脑皮层中活动的比喻，而旋律和面孔的例子不是。在后两者中，大脑结合了恒定表征和当前输入信息来做出详细预测。这一过程发生在大脑皮层的每一个区域，无处不在。是它令你能够对当前身处的房间进行具体预测；是它使你不仅能够预测别人要说的话，还能判断出他们的语气、口音，以及声音从何而来；是它使你准确知道脚何时触及地板以及爬楼梯时的感觉；是它使你能够用脚签名抑或是接住投来的球。

本章所探讨的3个大脑皮层记忆特征（存储序列、自-联想回忆和恒定表征），是在回忆的基础上预测未来的必要成分。在下一章中，我将探讨智能的本质，即预测。

第五章 智能理论的新框架

1986年4月的一天，我坐在办公室里，苦思“理解”究竟意味着什么。几个月来，我一直纠结于这一根本问题：如果大脑并不产生行为，那么它所做的是什么？当你在聆听演讲时，大脑做了什么？当你在阅读本书时，大脑在做什么？信息进入大脑，却不再出来，这中间发生了什么呢？你此刻的行为应该都是些基本行为，如呼吸和眼动，然而想必你已经了解，在你阅读和理解这段话的时候，大脑所作的努力要比这些基本行为多得多。理解一定是神经元活动的产物，但是，这些神经元在理解的时候，具体都做了些什么呢？

那天我在办公室里环顾四周，看到了熟悉的椅子、海报、窗子、植物、铅笔等等。有上百种物体和特征围绕着我。我的眼睛一瞥之下看到了它们，但“看到它们”这一事件并没有引发我的任何行为。尽管如此，我还是“理解了”这个房间以及里边的东西。我所做的正是赛尔的“中文屋”做不到的事情，而且还不需要通过墙壁缝隙传递任何信息。虽然没有任何行为来证明这一点，但我的确达成了理解。那么，这种“理解”究竟是什么意思呢？

就在我为此苦思冥想之际，脑中突然灵光一闪，原本的一团乱麻瞬间化作一片澄明。当时我只问了自己一个问题：如果一个我从未见过的新东西，比如一个蓝色的咖啡杯，突然出现在房间里，那么将会发生什么？

答案似乎很简单：我会注意到这个原本不属于这里的东西。作为新的事物，它会吸引我的注意力。我并不需要有意识地问自己这个咖啡杯是不是新出现的，它自己就会从环境中跳出来。在这看似微不足道的答

案背后，潜藏着一个强大的概念。要注意到某件东西发生了变化，我大脑中某些原本不活跃的神经元将会变得活跃起来。然而这些神经元如何知道蓝色咖啡杯是新出现的，而房间里其他上百件东西不是呢？对这个问题的答案至今仍然让我感到吃惊：我们的大脑利用记忆不断地在对我们所看、所听和所感的一切进行着预测。当我环视房间时，我的大脑利用记忆，在我体验事物之前就形成了对我所将要体验的事物的预测。绝大多数的预测是在意识不到的情况下发生的。就像是大脑的不同部位不停在自问自答："电脑在桌子中间吗？是的。它是黑色的吗？是的。台灯在桌子的右上角吗？是的。字典还在我放的位置吗？是的。窗子是长方形的，墙是垂直的吗？是的。在这个时间段，阳光照进来的方向正确吗？是的。"然而，当环境中出现了某些我没有记忆过的视觉模式时，预测就被打乱了，而我的注意力就会被吸引到这个错误上。

当然，大脑在作出预测时并不会同自身交流，也不会以序列的方式进行预测。它也不会对咖啡杯这类新出现的物体进行预测。你的大脑以一种并行的方式，对构成我们所生活的世界的每一处不断地进行预测。它能够很容易地发现一个奇怪的结构，一个变形的鼻子，或一个异常的动作。这种下意识的预测无时无刻地发生着，然而即使认识到这一事实，我们也无法立刻意识到它，这也许正是它的重要性被长久忽视的原因。它们发生得是如此轻松自然，以至于我们无法想明白大脑里面到底发生了什么。我希望这一强大的想法能给你留下深刻的印象：预测无时无刻不在。我们所谓的"感知"——也就是世界在我们眼中的样子——并不只是基于我们的感觉。我们所感知到的，是我们的感觉和源于大脑记忆的预测两者之结合。

* * *

几分钟后，我设计了一个思维实验，来说明我在那一刻所理解到的事情。我称之为"改变了的门"，它是这样的：每天当你回到家时，通常会用几秒钟的时间穿过大门。你伸出手转动把手，走进去，然后在身后

关上门。这是一个牢牢确立了的习惯，你经常做，但却很少注意到它。假如当你外出时，我溜进你家，把你的门稍加改动了一些。这一改动可以是任何事情，比如我可以将把手往旁边移动几厘米，或者将圆形把手换成门闩，或将从黄铜把手换成镀铬。我还可以改变门的重量，将中空门换成实心橡木门或者相反。我可以将铰链弄紧令其吱吱作响，或使它们润滑无摩擦。我可以扩大或缩小门及框架，改变它的颜色，在猫眼的位置安一个门环，或者加一扇窗。我能想象出1000种你所不知道的变化来改造你的门。当你回到家，准备打开门时，你会很快发现有什么地方不对劲。你可能需要几秒钟才能意识到究竟哪里不对，但你会立刻注意到发生了变化。伸手去摸被移动了的把手时，你会意识到，它没在原来的位置上。或者，看到门上的新窗子时，你会觉得怪怪的。如果改变了门的重量，你会以错误的力量推动它，然后感到很惊讶。问题的关键点是，你会在极短的时间内，注意到这上千种变化中的任何一种。

你是怎么做到的？又是如何注意到这些变化的呢？如果让人工智能或计算机工程师来解决这一问题，他们会首先创建一个有关门的所有特性的列表，并将它们输入数据库，包括门的所有属性字段，并没有你家大门的特定条目。当你走到家门口时，计算机会查询整个数据库，寻找你家大门的宽度、颜色、大小、把手位置、重量、声音等等信息。乍听起来，这同我形容环视办公室时大脑检查它的无数预测的过程很相似，但其实它们之间的区别是真实存在且影响深远的。人工智能的解决办法并不合理。首先，提前列举一扇门的所有属性是不可能的，那将会是一份长得无穷无尽的清单。其次，我们还需要对一生中的每个时刻所遇到的物体都列出一个清单。第三，就我们已有的有关大脑和神经元的知识来看，没有任何一点说明它们的工作方式是这样的。最后，神经元的速度太慢，无法实现计算机类型的数据库操作，如果是用它的话，你可能要花费20分钟而不是2秒才能注意到大门的变化。

因此，你对改变了的门的快速反应只存在一种解释：你的大脑对于每个特定时刻将要看到、听到和感觉到的事物进行了低层级的感官预

测，这一过程是并行发生的，大脑皮层的所有区域都在同时预测着它们接下来的体验。视觉区域对边缘、形状、物体、位置和运动作出预测；听觉区域对音调、声音源方向和模式进行预测；体感区域对触摸、质地、轮廓和温度进行预测。

“预测”意味着，参与知觉你家大门的神经元，在实际接收到感觉输入之前就开始活跃。当感觉输入真正到达时，再与所预期的相比对。当你走近家门，你的大脑皮层根据以往的经验形成了一系列的预测。当你伸出手来，它会预测你手指上的感觉，判断你什么时候会感知到门，以及当你触摸到门时，关节会是何种角度。当你开始推门时，你的大脑皮层会预测门的阻力会有多大，门会发出怎样的声音。当这些预测都与实际吻合时，你就会若无其事地穿过大门，根本意识不到对这些预测进行了验证。但是，如果你对门的预期被实际情况打破了，那么这种预期错误就会引起你的注意。正确的预测导向了理解——大门与平日无异。不正确的预测则会导致混乱，并引起你的注意——门闩的位置不对，门变轻了，门没在中间，把手的材质变了，等等。我们的所有感官都在以并行的方式不断进行着低层级的预测。

但这些还不是全部，我还要提出一个更为强大的理论——预测不只是你大脑的功能之一，它更是整个大脑皮层的主要功能和智能的基础。大脑皮层是一个专司预测的器官。如果想要了解什么是智能，什么是创造力，大脑如何工作以及如何构建智能机器，我们必须了解这些预测的本质，并弄清大脑皮层如何形成它们。即便是行为，也最好是作为预测的副产品来加以理解。

* * *

我不太清楚第一个提出预测是理解智能的关键的人是谁。在科学和工业界没有任何人发明过全新的东西。人们看到的只是如何将现有的想法融入新的框架。在被发现之前，新想法的成分通常是浮动于科学话语圈周围，通常情况下的新发明就是将这些成分组合成一个有机的整体。

同样，大脑皮层的主要功能就是预测这一观点也不是全新的。它已经在许多时候以多种形式存在了很久。但在大脑理论和智能定义方面，它并未得到应有的中心地位。

具有讽刺意味的是，一些人工智能的先驱们曾有过类似的概念，希望用计算机建立一个世界的模型，并用它来进行预测。例如在1956年，D.M.麦凯（D.M.Mackay）就曾提出，智能机器应该有一个“内部反应机制”，来“匹配所接收到的信息”。他没有使用“记忆”、“预测”这样的字眼，但他的思路是一样的。

20世纪90年代中期以来，如“推理”、“生成模型”、“预测”等名词悄然进入了科学术语表。它们所指的都是相关的想法。纽约大学医学院的鲁道夫·李纳斯（Rodolfo Llinas）在2001年出版的著作《旋涡中的“我”》（*i of the vortex*）中写道：“预测未来事件结果的能力对于成功的运动至关重要，并且极有可能是大脑最根本和最普遍的功能。”布朗大学的大卫·芒福德（David Mumford）、华盛顿大学的拉杰什·拉奥（Rajesh Rao）以及波士顿大学的斯蒂芬·格罗斯伯格（Stephen Grossberg）等众多科学家，也都曾以各种方式总结过反馈和预测的作用，并将之理论化。数学领域专门有一个子领域，致力于贝叶斯网络的研究，该网络以统计学先驱英国人托马斯·贝叶斯（Thomas Bayes，生于1702年）命名，使用概率论来进行预测。

我们目前所缺乏的，正是将这些分散的点点星光组织起来，形成一个统一的理论框架。这一点我认为还没有人做到，也正是这本书想要达到的目标。

* * *

在深入探讨大脑皮层如何进行预测之前，让我们再多看一些例子。对这个问题思考得越多，你就会越发认识到，预测无处不在，它是你理解世界的基础。

今天早餐我做了薄煎饼。在此过程中的某一刻，我伸手到厨台下去开柜门。凭着直觉，我不用去看就知道会摸到什么——这应该是柜门把手——也知道什么时候会摸到它。我拧牛奶瓶盖，知道它会转动并打开。我放上煎锅打开开关，预期旋钮被轻轻推入，然后稍微费力旋开，大约一秒钟后，火苗会“噗”的一声冒出来。在厨房里的每一分钟，我都会做出几十或上百个动作，每一个动作都会包含许多预测。之所以能肯定这一点，是因为如果在这些常见的动作中有任何一个与预期不同，我都会立刻注意到它。

在你走路时，每次落脚，你的大脑会预测你的脚何时停止移动，以及脚下的东西有多少弹性。当在楼梯上一脚踏空时，你会立刻意识到出错了。在落下的脚超出预期中台阶高度的一刹那，你就明白自己有麻烦了。你的脚没有任何感觉，而你的大脑却预测此时应该踩到台阶了，此时这个预测与实际情况便是不相符的。一个由计算机驱动的机器人会毫无知觉地摔下来，根本意识不到有什么不对劲。然而一旦你的脚超出大脑预期应该停止的地方哪怕是几厘米，你就会马上意识到。

当你听一首熟悉的旋律时，你会在头脑中提前听到下一个音符；在你听自己最喜欢的专辑时，下一首歌开始之前几秒钟你就能听到它的前奏。这是怎么回事呢？原来，在你听到下一个音符时大脑中应该激活的神经元，在你实际听到它之前就提前激活了，于是你在头脑中“听”到了这首歌。神经元的激活是对记忆的反应。这种记忆的持续时间长得惊人。在距离第一次听某张音乐专辑的多年以后，当再次听到它时，仍能于一首歌曲结束后自然而然地提前听到下一首歌的前奏，这种现象非常常见。而当你随机播放一张你所中意的CD时，就会产生一种由轻微的不确定而引发的愉悦感，因为你知道自己对下一首歌曲的预测是错的。

听人说话时，你往往会在他们一句话说完之前就提前知道他们要说什么——至少你认为自己知道。有时候，我们所听到的甚至不是说话者实际说的，而是我们希望听到的。（小时候这种事曾多次发生在我身

上，以至于妈妈带我去看了两次医生，检查我的听力是否有毛病。）之所以会出现这种情况，部分原因是因为人们在谈话时总会使用一些常见的词语和表达方式。如果我说："人之初，性本……"，你的大脑就会在我说出"善"字之前，激活代表它的神经元（而如果你不熟悉这句三字经的话，可能就不知道我在说什么）。当然，我们并不是总能知道别人要说什么，预测也并不总是准确的。确切地说，我们的大脑是通过对即将发生的事情进行概率预测来工作的。有时候我们能够确切知晓将要发生的事情，而另一些时候我们的预期则分布于几种可能性之中。假设我们在一家小餐馆的桌子旁吃饭，我对你说："请递给我……"，无论我接下来说的是"盐"、"辣椒"还是"芥末"，你的大脑都不会感到惊讶。从某种意义上说，你的大脑已经立刻预测到所有这些可能的结果了。但如果我说："请递给我人行道。"你就会觉得有什么不对劲了。

再回到音乐上来，在这里我们同样可以看到概率预测。如果给你听一首从没听过的歌，你仍然可以有相当强的预期。在西方音乐中，我会预期听到一个规则的节拍、一个重复的节奏，我会预期包含相同数量小节的乐句，并期望歌曲在主音高上结束。你可能不明白这些术语的意思，但是——假设你听过类似的音乐——你的大脑会自动预测节拍、重复的节奏、乐句以及歌曲的结束。如果一首曲子违反了这些原则，你马上就会意识到有问题。花一秒钟想想这其中的奥秘吧。在听一首从未曾听过的歌时，你的大脑所经历的模式是它从未经历过的，然而你仍然可以作出预测并判断是否有哪里不对。这些下意识的预测是以存储在你大脑皮层里的一套记忆为基础的。你的大脑不能确切地说出接下来会发生什么，但它仍然能够预测哪些音符模式很可能出现，而哪些不可能。

我们都有过这样的经历：突然注意到持续的背景噪声——如远方的气锤声或单调的背景音乐——停了下来，而在它们发出持续的声音时我们根本不曾注意到。你的听觉区域预测的是音乐一秒接一秒的连续性，只要声音没有变化，你就不会留意。而如果声音突然停止，就会破坏你的预测，从而吸引你的注意力。有一个历史上真实存在的例子：就在纽

约市刚刚停运高架列车的那段时间，经常有人半夜打电话报警，声称被什么东西惊醒了，而报警的时间，往往是以前列车经过他们公寓的时间。

人们喜欢说“眼见为实”。然而，在我们所看到的事物中，实际看到的与期望看到的，在比例上旗鼓相当。这方面一个最有趣的例子与研究人员称为“填补”的现象有关。你可能已经知道，在每只眼睛的视网膜上都有一个盲点，这里是视神经穿过叫作“视盘”的小洞离开视网膜的地方。这个区域没有任何光感受器，因此在视野中相应的那一点上，可以说是永久失明的。但是你通常不会注意到它们。这是为什么呢？原因有两个，一个很平常，另一个则富有启发性。平常的那个原因就是，你的两个盲点并不交叠，因此一只眼睛能够补偿另外一只。

但有趣的是，即使你只睁一只眼，仍然不会注意到盲点。你的视觉系统“填补”了缺失的信息。当你闭上一只眼睛，观看风格华丽的土耳其编织地毯或是樱桃木桌面上的波状纹理时，你不会看到一个黑洞。地毯上全部的线结和木头纹理上所有的暗结都会在被盲点覆盖的时候消失在视网膜的图像里，然而你所看到的却是无缝延续的纹理和颜色。你的视觉皮层利用记忆中类似的模式，形成了一个连续的预测信息流，用来填补任何缺失的输入信息。

这种“填补”发生在视觉图像的各个部分，不只是在盲点上。比方说，我给你看一张海岸的照片，一棵漂流木倒在几块海边的岩石上，岩石和木头之间的边界清晰可辨。然而，如果我们将图像放大，你就会看到岩石和漂流木相接之处的纹理和颜色非常相似。在放大的视图中，木头的边缘很难与岩石相区分。如果我们观看整个场景，漂流木的边缘就很清楚，然而实际上这个清楚的边缘是我们从图像的其他部分推断出来的。当我们看世界时，我们感知到不同物体被清晰的线条和边界分开来，但进入我们眼睛的原始数据通常是纷乱而模糊的。我们的大脑皮层将它认为应该位于那儿的信息填补到那些缺失或混乱的部分，于是我们

便看到了清晰的图像。

视觉的预测同时也是眼动方式的一个功能。在第三章中我曾提到过扫视，即是说，每秒钟大约有3次，你的眼睛会从一个注视点上突然跳到另一个。一般来说你不会意识到这些眼动，而且你通常不会有意识地去控制它们。每当你的眼睛跳到新的注视点时，进入大脑的模式都会完全不同于上一个注视点，因此，你的大脑在一秒内会看到3次完全不同的东西。扫视并不完全是随机的。当你看到一张脸时，通常会先注视一只眼睛，然后是另一只，左右不停地来回看，偶尔注视一下鼻子、嘴、耳朵以及其他特征。你感知到的只是“脸”，而你的眼睛注视的却是眼睛——眼睛——鼻子——嘴——眼睛，等等。我知道你所体验到的并非如此。你所感知到的是这个世界的连续景象，但进入你大脑的原始数据，抖得就像出自一台用废了的摄像机。

现在来想象一下，你遇到了一个在长眼睛的地方长出一只鼻子的人。你的眼睛会首先注视在他的一只眼睛上，然后扫视到另一只眼睛，但你在那里没有看到眼睛，而是看到了一只鼻子。你肯定会觉得有问题。而你之所以会有这个感觉，是因为你的大脑对于将要看到的事物有一个期望或者说是预测。当你预测看到眼睛，结果却看到了鼻子，这个预测就被打破了。因此，伴随着每一次扫视，你的大脑每秒钟都会对接下来将要发生的事情作出好几次预测。一旦预测出错，就会马上引起你的注意。这就是为什么当我们遇到身体有畸形的人时，很难忍住不看他们。如果遇到长了两个鼻子的人，你能忍住不盯着看吗？当然，如果与这个人一起生活一段时间，你就会习惯他的两个鼻子，不会再将它当作非同寻常的事物去加以关注了。

现在来想想你自己。你正在做着什么样的预测？当翻动这本书时，你会预期页面稍稍弯曲，并以一种可预见的方式翻动，它应该与翻动封面的方式是不同的。如果你正坐在椅子上，你会预测身体所感受到的压力是持续不变的。但如果座位开始变湿，或者向后滑，或者出现了任何

一种意料之外的变化，你就会将注意力从书上移开，去弄清楚发生了什么事。花点时间观察自己，你会明白，你对世界的看法和理解也同预测紧密相连。你的大脑建立了关于这个世界的模型，并不断地将它与现实相比照，得益于这个模型的有效性，你才能够知道自己在哪里以及正在做什么。

预测并不仅仅局限于“看”和“听”这些低级的感觉信息模式。到目前为止，我的讨论一直局限于这一类例子，因为它们是介绍智能新框架的最简单方式。然而，根据蒙卡斯尔的理论，适用于低级感觉区的原理也必然适用于所有皮层区域。人类的大脑之所以比其他动物的大脑更聪明，是因为它可以对更抽象的模式和更长的时间模式序列作出预测。要预测我妻子今天看到我时会说什么，我必须知道她曾经说过什么（今天是周五，垃圾桶必须在周五晚上放到马路边上，上周五我没有按时做这事），还有她脸上当时的表情。当她张开嘴时，我有非常强烈的预感她会说什么。在这种情况下，尽管我不知道她确切的话语会是什么，但我的确知道她会提醒我把垃圾桶拿出去。最重要的一点是，较高的智能并不是一种有别于感知智能的过程，它从根本上同样有赖于大脑皮层记忆和预测算法。

请注意，我们所使用的智力测验，在本质上就是对预测能力的测验。从幼儿园开始直到大学，智力测试的基础都是做出预测。比如，给你一个数字序列，问你下一个数字应该是什么；或者给你看一个复杂物体的3个面，让你判断下面哪一个图形也是这一物体的一个面；或者问，字母A相对于字母B，就像字母C相对于字母__？

科学本身就是一种预测训练。我们通过一套假设和测试的过程，来推进我们对世界的认识。这本书本质上也是对于智能是什么以及大脑如何工作的预测。甚至连产品设计从根本上说也是一种预测的过程。无论是设计服装还是手机，设计师和工程师们都会试图预测竞争对手会做什么，消费者想要什么，新的设计将花费多少以及什么样的风格会受欢

迎。

智能是以对世界中模式的记忆和预测能力来衡量的，这些模式包括语言、数学、物体的物理属性以及社会环境。你的大脑从外界接收模式，将它们存储成记忆，然后结合它们曾经的情况和正在发生的事情来进行预测。

* * *

看到这里，你可能会这样想：“即使躺在黑暗中什么也不做，我也是有智能的，因为我的大脑在进行预测——这个说法我可以接受。正如你所说，我并不需要为了理解或得到智能而去做什么，但这种情况难道不是一种例外吗？难道你真的认为有智能的理解与行为是完全无关的吗？最终让我们拥有智能的不正是行为而非预测吗？因为毕竟，行为才是生存的最终决定因素。”

这个问题问得很好。当然，对于动物的生存来说，最重要的因素还是行为。预测和行为并不是完全分离的，但它们之间的关系比较微妙。首先，大脑皮层是在动物进化出复杂的行为之后才被进化出来的，因此，大脑皮层的存在价值，必须首先从它对动物的已有行为所提供的渐进改进方面加以理解。先有行为，然后才有智能。其次，我们所感知到的大多数事物，在很大程度上都依赖于我们的行为和在世界上的运动。因此，预测和行为密切相关。让我们来看看相关的例子。

哺乳动物进化出了很大的新大脑皮层，因为大脑皮层能够带给它们一定的生存优势，而这种优势最终是根植于行为的。但在最初，大脑皮层只是有助于更有效地利用现有的行为，而不是创造出全新的行为。要弄清楚这一点，我们需要回顾一下大脑的演化历程。

几亿年前，在多细胞生物开始出现在地球的各个角落之后不久，出现了简单的神经系统，然而真正的智能是随着我们的爬行动物祖先出现

的。爬行动物成功地征服了大陆，它们遍布各个大洲，并分化出了无数物种。它们拥有敏锐的感官和能够赋予它们复杂行为的发达大脑。它们的直系后裔——今天幸存的爬行动物们，仍然具有这些特点。比如短吻鳄，它拥有和你我一样的复杂感官——发达的眼睛、耳朵、鼻子、嘴和皮肤；它还拥有实现复杂行为的能力，包括游泳、奔跑、隐藏、捕猎、伏击、晒太阳、筑巢和交配。

人类大脑和爬行动物大脑之间的差别是什么呢？这一差别既大又小。说小是因为，粗略看来，爬行动物大脑中有的东西，人脑中都有；说大是因为，人脑中有一些爬行动物大脑所没有的真正重要的东西，也就是大脑皮层。有时你会听到人们提及“旧脑”或“原始脑”，每个人的脑中都有这些较为古老的结构，就像爬行动物一样。它们调节着血压、饥饿感、性欲、情绪以及运动的许多方面。例如，当你站立、保持平衡或行走时，你主要依靠旧脑。如果你听到一个可怕的声音，感到恐慌并开始逃跑，那也是你的旧脑在起作用。许多有趣且有用的事情，只要有爬行动物那样的大脑就足够胜任了。既然看、听和动作都不需要新大脑皮层的参与，那么它究竟是做什么用的呢？

哺乳动物比爬行动物更聪明，靠的正是它们的新大脑皮层（neocortex，源于拉丁语，意为“新树皮”或“新外皮”，因为大脑皮层确实将旧脑覆盖了起来）。几千万年以前，大脑皮层首先出现在哺乳动物脑中。几百万年前，人类的大脑皮层面积急剧扩大，使得人类比其他哺乳动物更聪明。不要忘了，大脑皮层由一个常见的重复元素构建而成，人类大脑皮质与我们的哺乳动物亲戚们有着相同的厚度和结构。当进化快速地造就一些东西时——就像它对大脑皮层做的那样，是通过复制已有结构来实现的。通过为一种通用皮层算法加入更多的元素，我们获得了智能。有一种常见的误解，认为人类大脑是耗时数十亿年进化出的巅峰之作。如果考虑到整个神经系统的话，可能的确如此。然而，大脑皮层本身也是一个相对较新的结构，并没有足够长的时间去经历长期进化的精制打磨。

如何理解新大脑皮层？为何记忆和预测是解开智能之谜的钥匙？接下来我就要谈到这些问题的核心。让我们从没有大脑皮层的爬行动物大脑开始说起吧。人们发现进化中，如果为原始脑的感觉通道加上一个记忆系统（即大脑皮层）的话，动物就能获得预测未来的能力。想象一下，让古老的爬行动物大脑仍然执行原有功能，但现在的感觉模式会同时输入到新大脑皮层中，并在记忆系统中存储起来。在未来的某个时间，当动物遇到相同或类似的情况时，记忆系统就会识别出相似的输入模式，并回忆起过去情况下发生的事情。被提取出来的记忆将与感觉输入流进行比对，它既"填补"了当前的输入信息，又预测了接下来的情况。通过对实际的感官输入和回忆的对比，动物不仅可以了解它当下的处境，还可以预见未来。

现在让我们设想，大脑皮层不仅能记住动物已经见过的东西，而且还记得在相似情形下旧脑所作出的行为反应。我们甚至不必假设大脑皮层能够区分感觉和行为，对它来说，这两者都只是模式。当我们的动物发现自己处在相同或类似的情形中时，它不仅能预见未来，而且能够回忆起采取何种行为模式能让这个未来场景变为现实。如此一来，记忆和预测便使得动物能够更聪明地利用它已经存在的（旧脑）行为。

假如你是一只第一次学走迷宫的老鼠，因不确定性和饥饿而备受折磨，你就会利用旧脑固有的技能在新环境中摸索——去听、去看、去嗅，溜着墙壁向前爬。所有这些感觉信息都被旧脑使用，但也会传输到你的大脑皮质，在那里存储起来。在将来的某个时候，当你发现自己身处相同的迷宫时，你的大脑皮质会认出当前的输入与曾经看到过的相同，并提取出它所存储的表征着过去情形的模式。从本质上说，它可以令你预见到不久后的未来。如果你是只会说话的老鼠，这时你可能会说："哦哦，我认得这个迷宫，我记得这个角落！"当你的大脑皮层回忆起过去的情形时，你的眼前就会浮现起上次在迷宫里见到的奶酪和你得到它的过程。"如果在这里右转的话，我记得接下来会发生什么。在走廊的尽头有一块奶酪，我好像已经能看到它的样子了！"当你在迷宫里

溜达时，你所做的动作，如抬腿、捋胡须等，都是依靠古老而原始的脑结构来完成的。而你的（相对较大的）新大脑皮层则能让你记住去过的地方，在未来再次认出它们，并预测接下来将要发生的事情。没有大脑皮层的蜥蜴对过去的记忆能力就差多了，迷宫对它来说每次都是新的。而你（作为一只老鼠）之所以能够了解世界并预知不远的将来，全是得益于大脑皮层的记忆功能。你能看到每一个决定所导向的奖励和危险的生动画面，因此可以更为有效地穿行于你所处的世界。你确实可以“看到”未来。

但是请注意，你并没有做出任何特别复杂的或是全新的行为。你并没有造出一架滑翔机，好带你飞往走廊尽头的奶酪。你的大脑皮层形成的感觉模式预测使你能够看到未来，但你可用的行为库却几乎不受影响。你的行为能力，像蹿行、攀爬和摸索等，仍同蜥蜴没有太大差别。

随着大脑皮层进化得越来越大，它所能够记忆的关于世界的信息也越来越多。它能够形成更多的记忆，作出更多的预测。而这些记忆和预测的复杂性也随之增加。然而除此之外，还发生了一些更加非凡的事情，由此造就了人类独特的智能行为能力。

人类的行为超越了老鼠所拥有的那些旧的基本技能。人类已将大脑皮层的进化带到了一个新的高度。只有人类创造出了书写和口头的语言，只有人类烹饪食物、缝制服装、开飞机、盖摩天大楼。我们的运动和规划能力远远超过了那些与我们最为亲近的动物亲戚。原本用来进行感觉预测的大脑皮层，是如何创造出人类特有的极其复杂的行为的呢？这种高级的行为又是如何突然进化出来的呢？对这个问题我有两种答案。一种是：大脑皮层的算法极为强大灵活，只需一点人类独有的重新调整，就可以创造出新的复杂行为。另一种答案是，行为和预测是同一件事物的两面，尽管大脑皮层能够预见未来，但它只有在对所执行的行为有所了解时，才能作出准确的感觉预测。

在老鼠找奶酪的例子里，老鼠记住了迷宫，并利用这一记忆来预测

自己会在拐角处看到奶酪。它可以向左转也可以向右转，只有同时记住奶酪和正确的行为——在交叉路口右转，才能让找到奶酪的预测成真。虽然这只是一个简单的例子，但它道出了感觉预测与行为紧密相关的要义。所有的行为都能改变我们所看到的、听到的和感觉到的。我们在每时每刻所得到的大部分感知又高度依赖于我们自身的行为。将手臂在眼前动一动，你的大脑皮层必须知道它发出了移动手臂的指令，才能作出“看见手臂在动”的预测。如果大脑皮层没有做出相应的运动指令，你却看到手臂在动，一定会感到无比惊讶。对此有一个最简单的解释：假设你的大脑先发出了移动手臂的指令，然后再预测会看到什么。我认为这种假设是错误的。相反，我认为大脑皮层先预测会看到手臂移动，然后这一预测引发它发出运动指令，从而使预测成为现实。你首先产生想法，然后采取行动来使想法成真。

现在，我们来看看使人类的行为库得以极大扩展的那些变化究竟是什么。猴子的大脑皮层和人类大脑皮层的是否存在着某种物理差异，可以用来解释为何只有人类拥有语言和其他复杂的行为呢？人类大脑的体积大约是黑猩猩的3倍，但这也并不仅仅是“越大越好”的问题。理解人类行为飞速发展的关键，存在于大脑皮层各个区域和旧脑各部分的神经连接上，简单来说就是，我们大脑的神经连接比较不一样。

让我们来仔细看一下。每个人都知道，大脑分为左、右半球。但还有另一种不太为人熟知的划分，而它正是我们寻找人类差异所需要的。所有的大脑，尤其是体积较大的，都将大脑皮层分为前、后两个部分。科学家们使用“前部（anterior）”和“后部（posterior）”来分别表示。分离前部和后部的是一个大的裂隙，称为中央沟（central sulcus）。大脑皮层的后部接纳从眼睛、耳朵和触觉传来的感觉输入，是大部分感知觉产生的地方。前部皮层则包含涉及了高级计划和思考能力的区域，还有主要负责肌肉运动并产生行为的运动皮层。

随着灵长类动物的大脑皮层进化得越来越大，前部更是超出了原有

的比例，特别是人类的大脑皮层。与其他灵长类动物和早期人类相比，我们的额头变得很大，方便用来装载极大的前部大脑皮层。但是这种皮层的增大并不足以解释与其他动物相比我们在运动能力上的改进。我们之所以拥有进行复杂运动的能力，是由于我们的大脑运动皮层与身体的肌肉之间存在着更多的连接。在其他哺乳动物身上，前部大脑皮层在运动行为中并没有起到直接的作用，它们在很大程度上还是依靠旧脑的部分来产生行为。与此相反，人类的大脑皮层在运动控制上取代了大脑的其他部分。如果一只老鼠的运动皮层受到损坏，它可能不会表现出明显的缺陷；而如果损坏了一个人的运动皮层，他／她就会瘫痪。

人们经常向我问起，海豚的大脑难道还不够大吗？是的，海豚有一个很大的大脑皮层，虽然海豚的大脑皮层结构比起人类较为简单（只有3层，人类有6层），但以其他尺度去衡量，已经算很大了。海豚很可能能够记住并理解很多事情，它可以认出自己的同类，对自己的生活可能还拥有良好的自传体记忆。它可能记得海洋中曾到过的每一个角落和缝隙。然而，尽管它们表现出了一些很复杂的行为，但仍无法企及人类。由此我们可以推测，它们的大脑皮层对行为的影响并不是决定性的。大脑皮层的进化主要是为了提供对外部世界的记忆，拥有大的大脑皮层的动物，可以和你我一样感知这个世界。然而人类的独特之处在于，人的大脑皮层在产生和控制行为方面起着主导和超前的作用。这就是为什么我们拥有复杂的语言和工具，而其他动物没有的原因，也解释了为什么我们可以写小说、上网冲浪、发送探测器到火星和建造邮轮。

现在，我们已经能够看到一幅完整的画面。大自然首先创造出一些动物，如爬行动物，它们拥有复杂的感官和复杂但有限的行为模式。随后大自然发现，通过给这些动物增加一个记忆系统，并输入感觉信息流，这些动物就可以记住过去的经验。当动物身处相同或类似的情境下时，过去的记忆就会被唤醒，引发对接下来可能会发生的事情的预测。因此，当记忆系统将预测反馈给感觉信息流的时候，智能和理解就出现了。这些预测便是理解的本质。理解一件事情，就意味着你能够对它作

出预测。

大脑皮层是在两个方向上进化的。首先，在可存储的记忆形式上向更大、更复杂的方向进化，这样它便可以记住更多的东西，并能以更复杂的关系为基础进行预测。其次，它开始与旧脑的运动系统进行交互。要预测将要听到、看到和感觉到的事物，大脑皮层需要知道正在进行的动作是什么。人类的大脑皮层接管了我们大部分的运动行为。它不仅仅基于旧脑的行为作出预测，还指导行为的发生以实现它的预测。

人类的大脑皮层非常之大，因此有着相当庞大的记忆容量。它能够不断地预测你将要看到、听到和感觉到的东西，而这些大多发生在你毫无意识的情况下。这些预测就是我们的思想，当它们与感觉输入结合之后，就形成了我们的知觉。我将这个对于大脑的观点，称为智能的"记忆-预测框架"。

如果塞尔的"中文屋"包含一个类似的记忆系统，能够预测下一个汉字是什么以及故事的下一步会发生什么，我们就可以满怀信心地说，这个房间懂中文，也理解这个故事。到这儿，我们可以看出阿兰·图灵错在了哪里——智能的证据是预测，而不是行为。

现在我们已经作好了准备，来深入探讨有关大脑的记忆-预测框架的新理论。要对未来事件作出预测，你的大脑皮层必须存储模式序列。要唤起合适的记忆，它必须根据新旧模式之间的相似性来检索模式（自-联想记忆）。最后，记忆还必须以恒定的形式存储，这样才能便于将过去事件的知识应用到相似但不完全相同的新情境中去。大脑皮层在物理层面上是如何完成这些任务的？它的层级结构具体又是怎样组织的？这些将是下一章的主题。

第六章 大脑皮层是如何工作的

试图理解大脑的工作方式，就像做一个巨大的拼图游戏。你可以采取两种途径完成拼图。如果采用“自上而下”的途径，那么你首先要对整个拼图的大致样子有概念，然后以此为依据决定忽略掉哪块拼板，搜索哪块拼板。如果采用“自底向上”的途径，你就要集中精力在每一块拼板上。你要研究他们独特的特征，寻找互相匹配的拼板。如果你对整个拼图没有概念，那么自底向上的途径往往就是唯一的处理方式。

“理解人脑”的拼图游戏尤其令人生畏。由于缺乏一个较好的框架来理解智能，科学家们不得不采用自底向上的途径。但是，对于复杂如人脑这样的拼图，要完成它即使有一定的可能，那也是极其困难的。为了感受这个问题的难度，让我们想象一个由几千块拼板组成的拼图游戏。许多块拼板可以通过多种方式解释，就像每一块拼板会在两面各有一个图像，而其中只有一个是对的。每块拼板都奇形怪状，你很难判断两块拼板是否可以互相拼接。很多块拼板将不会在最终的拼图中出现，但你很难判断是哪些，以及会有多少块这样的拼板。每个月都有新的拼板邮寄给你。一些新的拼板会替代旧的拼板，就好像拼图游戏制作者在说：“我知道你已经在旧的拼图上工作了几年，但它们有些错误。对不起。请用这些新的拼板并等待我们的进一步通知”。不幸的是，你不知道最终结果到底会怎样，更糟的情况是，你可能知道一些，但它们是错的。

这个拼图的比喻可以很好地描述我们在建立关于大脑皮层和智能的理论时所遇到的困难。科学家们一百多年来收集的生物学和行为学数据是这个拼图游戏的碎片。每个月都有新论文发表，产生新的拼图碎片。有时来自一个科学家的数据和另外的数据有矛盾。因为这些数据可以有

不同的解释，因此几乎处处存在不一致。由于没有一个自上而下的框架，因此在要找什么，什么是最重要的、以及如何解释堆积如山的数据信息等方面毫无共识。通过自底向上的途径来理解人脑已经途穷。我们需要一种自上而下的框架。

记忆预测模型就扮演了这样的角色。它可以告诉我们如何开始将碎片拼在一起。为了进行预测，你的大脑皮层需要某种方式来记忆和存储关于一系列事件的知识。为了预测新出现的事件，大脑皮层必须形成恒定表征（invariant representations）。你的大脑需要创建并存储一个关于这个世界的模型，它独立于你所处的不同环境。知道大脑皮层必须做什么，可以指导我们理解它的架构，特别是它的多层设计和六层结构。

我们将要探讨的新框架，是第一次在这里介绍，我将要深入的细节可能对一些读者来说是个挑战。很多概念对你来说可能会很陌生，甚至对神经科学家们也是如此。但通过一定的努力，我相信任何人都能对这个新框架有基本的认识。这本书的第七章和第八章将会探讨该理论更广泛的影响，技术味道会淡一些。

我们的拼图之旅现在变成了寻找记忆预测假设的生物学支持。知道哪些碎片是与最终图像最相关的以后，就先不用去管其他大部分碎片。我们一旦知道了要找什么，这个任务就变得可控了。

同时我想强调，这个新的框架尚待完善。有很多事情我还不能理解。但是基于演绎推理、来自不同实验室的实验以及解剖学等，我已经理解了很多事情。在过去的5到10年间，来自神经科学的很多领域的研究者已经探索了一些想法，虽然这些想法使用了不同的术语，据我所知也没有人将这些想法形成一个总体框架，但这些想法与我的很类似。它们也会探讨自上而下和自底向上的处理，探讨人脑感知区域之间的模式传播，探讨恒定表征如何扮演了重要的角色。例如，来自加州理工学院的神经学家加布里埃尔·克赖曼（Gabriel Kreiman）、克里斯托弗·科克（Christof Koch）和来自UCLA的神经外科医生伊扎克·弗里德（Itzhak

Fried）发现了一些细胞，每当人们看到比尔·克林顿的照片时，这些细胞就会被激活。我的目标之一就是解释这些“比尔·克林顿”细胞是如何产生的。当然，所有理论都需要作出能在实验室中验证的预测。我在附录中建议了很多这类预测。既然我们知道要找什么，这个复杂系统看上去就没那么复杂了。

在本章的接下来的几部分，我们将逐层深入探讨大脑皮层的记忆预测模型是如何工作的。我们将首先探讨大脑新皮质的宏观结构和功能，然后逐步理解其中的每一部分，以及它们是如何形成一个整体的。

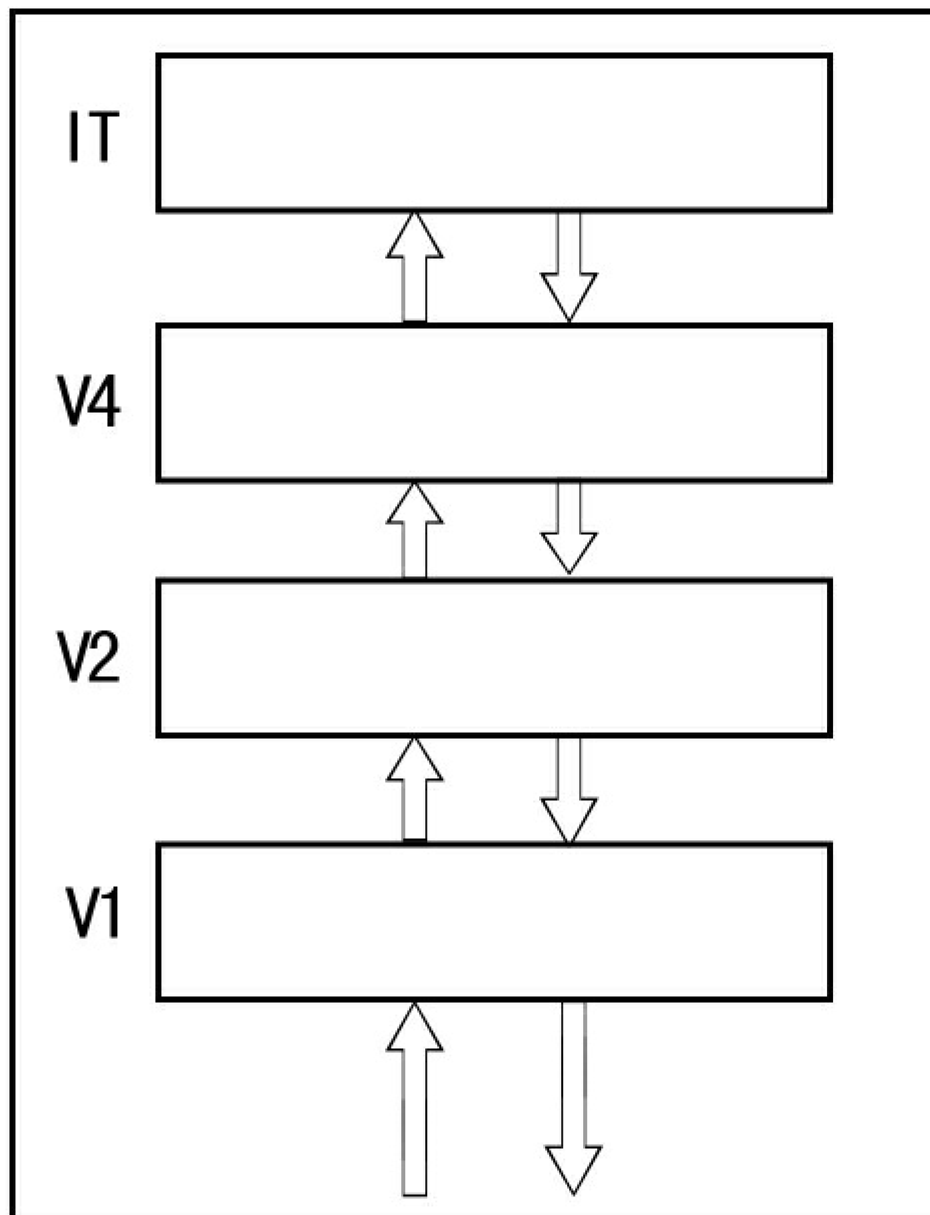


图1　识别物体的前4个视觉区域

恒定表征（Invariant Representations）

前面我将大脑皮层描述为餐巾那么大的一层细胞，有6张名片那么厚，由不同区域间的连接从整体上形成层次的结构。现在我想重新描述大脑皮层，来强调它的层次连接性。想象一下，我们将大脑皮层这张“餐巾”划分为不同的功能区——专注完成特定任务的大脑皮层区域——然后像煎饼一样叠在一起。你从一侧看过去，就会得到如图1所示的层次结构。提醒一下，大脑皮层并不是真的看上去如此，但这样的想象可以帮助你了解信息是如何流动的。我已经展示过4层皮层区域，感知输入从最底层进入，然后从低层逐层流动到高层。需要注意的是，信息是双向流动的。

图1显示了识别物体的前4个视觉区域，这些物体可以是一只猫、一座教堂、你的母亲、中国的长城，等等。生物学家把它们标注为V1、V2、V4和IT。视觉输入用向上的箭头表示，始于你双眼的视网膜，从图1中的底部开始传递到V1区。这个输入表示随时间变化的模式，由大约100万的神经轴突组成的视觉神经传输。

我们前面介绍了时空模式，但这里需要再强调一下，因为我们接下来会不断提到。记住，你的大脑皮层是一大片组织，其中包含很多功能区域，能够完成特定的任务。这些区域通过大量的神经轴突和纤维连接在一起，并互相传递信息。每时每刻，都有一些神经纤维会产生电信号（我们称作动作电位或信号尖峰），其他神经纤维则保持沉默。在神经纤维束上的这些信号，表示了某种模式。当你的眼睛刚看到物体的那一瞬，到达V1区的信号表示空间模式，而当你的眼睛扫过物体的时候，到达V1区的信号则表示时序模式。

如前所述，你的眼睛会不停地快速移动并停下来，大约每秒3次，我们称眼睛移动为“扫视”，称眼睛停下来为“注视”。如果一个科学家为你装上一个跟踪眼睛移动的装置，你会惊讶地发现，与平稳的视觉感受不同，你的扫视动作非常快速而不平稳。图2（a）展示了人眼在看一张人脸的时候是如何移动的。需要注意，注视并不是随机发生的。设想你能看到从这个人的眼睛接收的信号到V1层的动作模式，你会发现，视觉皮层每秒能够看到好几次全新的模式。

你也许会认为：“好吧，不过那仍然是同一张脸啊，只是偏移了一下而已。”你的这个想法有一部分是对的，但远不是全部。你视网膜上的光感受器是不均匀分布的。感受器集中在视网膜的中央凹上，越向外围就变得越稀疏。与此相反，大脑皮层的细胞则是均匀分布的。这样的结果就是，从视网膜传送到初级视觉区V1的图像是高度扭曲的。你的眼睛注视的是鼻子还是眼睛，得到的视觉输入信息是非常不同的，这就好像是透过来回剧烈抖动的、扭曲的鱼眼镜头观察似的。然而当你看到这张脸的时候，它并不是扭曲的，而且也没有跳来跳去。大部分时候你甚至感受不到视网膜的视觉模式发生过改变，更不要说有多剧烈了。你只是看到了一张“脸”。（图2b展示了沙滩景色的在视网膜上的扭曲情况。）这又涉及我们在第四章提到的“恒定表征”之谜。你“感知”到的并不是V1看到的。你的大脑是如何知道它在看同一张人脸呢？为什么你会不知道视觉输入是变化的而且是扭曲的呢？

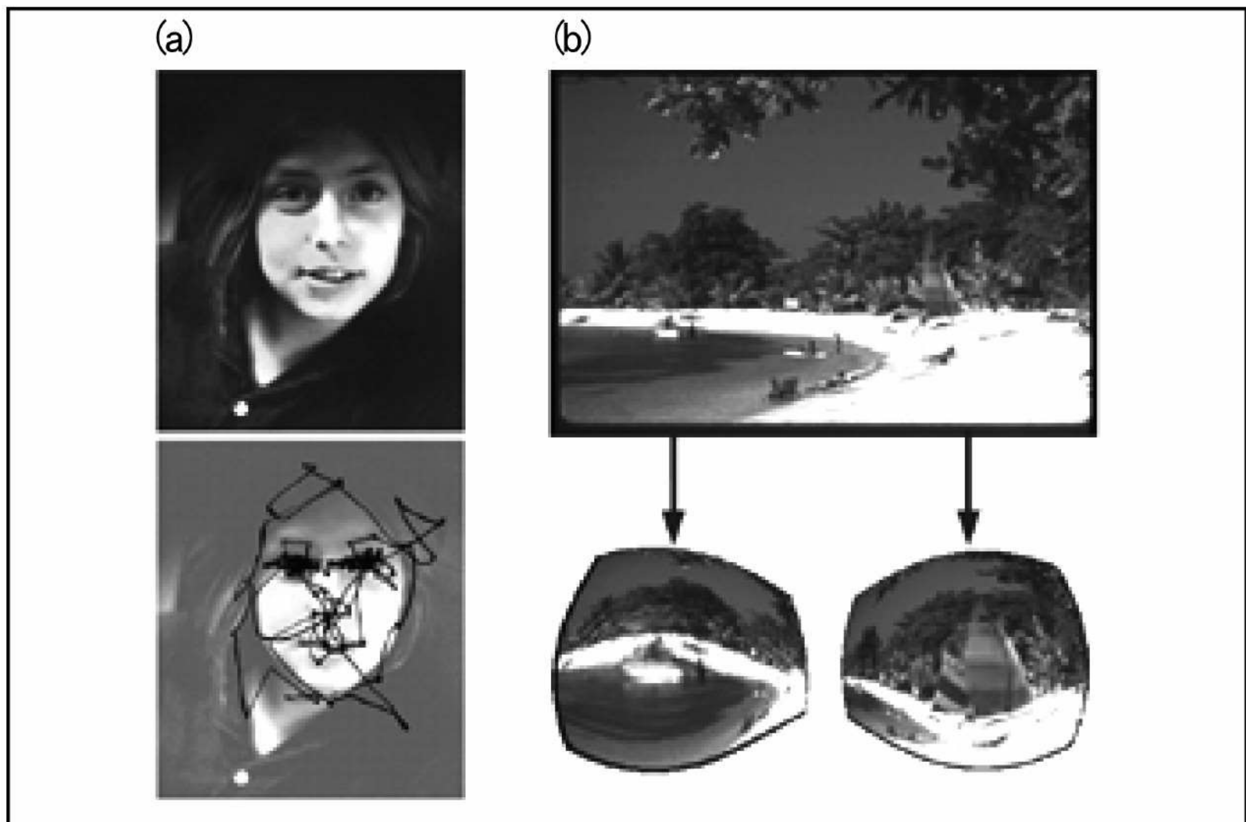


图2 （a）眼睛在看人脸的时候是如何移动的
（b）由视网膜视觉感受器的不均匀分布造成的扭曲

如果我们在V1区插一根探针来观察单个细胞的反应，我们会发现每个细胞只会对视网膜的一小块区域的视觉输入产生反应。这个实验已经被做过很多次，一直是视觉研究的核心问题。每个V1神经元都对应一个“感受野”，是整个视野中的一小部分，视野是指你眼前所能看到的全部。V1细胞并不拥有关于你看到物体的任何知识，如人脸、汽车、书籍以及其他物体等。它们只知道视野中针眼那么大小一丁点的区域的信息。

每个V1细胞也只会对特定输入模式类型产生反应。例如，当感受野中出现倾斜30度的线段或边界的时候，某种神经细胞就会被激发。这种边界本身并没有什么意义。它可能是某个物体组成部分，例如地板、远处棕榈树的树干、字母M的一侧，以及其他无限种可能的情况。伴随着每次注视，这个细胞的感受野都会变换到视野中新的区域。在一些注

视情况下，这个细胞会更加活跃，而其他情况下，这个细胞可能就不那么活跃。因此，每次你进行扫视，V1的很多细胞都可能会改变它们的活动。

然而，如果你把探针插在图1中最高的那个IT区，就会看到一些神奇的事情。我们会发现，当某个物体出现在视野中的时候，一些细胞会被激活并保持激活状态。例如，当一张人脸出现的时候，一些细胞就会变得活跃。只要你一直看着人脸，这些细胞就会一直保持活跃。它们的状态不会像V1细胞那样随着扫视而变化。这类IT细胞的感受野覆盖了大部分视野，只要看到人脸就会被激活。

让我们来揭开这个谜底。在从视网膜到IT区的4个不同层次的区域中，细胞从快速变化、空间相关、能识别细微特征的细胞，逐渐变成了稳定激活、空间无关、能识别物体的细胞。IT区细胞告诉我们我们的视野中出现了一张人脸。这种细胞通常被称为“人脸细胞”，只要有人脸就会被激活，不管这张脸是倾斜的，旋转的，还是部分被遮盖的。这是针对“人脸”的恒定表征。

这事儿说起来容易：通过4个阶段，瞧，我们识别出一张人脸来。但目前还没有计算机程序或数学公式能够像人脑那样鲁棒而普适地解决这个问题。然而我们既然知道人脑是通过有限步骤解决了这个问题，那么答案应该不难找到。本章的主要目的之一，就是解释识别一张人脸（如比尔·克林顿的脸）的细胞是如何产生的。在讲这些之前，我们还要先介绍一些基础知识。有很多轴突束会从IT等较高区域连回到V4、V2、V1等底层区域。更重要的是，这些反馈连接的数量至少有前馈连接一样多。

有相当长一段时间，大部分科学家都忽略了这些反馈连接。如果对大脑的研究仅仅是聚焦在大脑皮层如何对输入信息进行接收、处理和反应的话，是不需要研究反馈的。你只需要研究大脑皮层中从感知区到运动区的前馈连接就可以了。但是当你认识到大脑皮层的核心功能是进行

预测时，就需要将反馈考虑在内了，也就是说，大脑需要将信息送回到最初接收输入的区域。预测需要比较真正发生的事情和你预期发生的事情。真正发生的事情的信息会自下向上流动，而你预期发生事情的信息会自上向下流动。

同样的前馈-反馈过程发生在你所有感知皮层区域。图3展示表明，我们的视觉皮层结构与听觉和触觉是类似的。该图还显示有一些高层皮层区域，作为联合区，它们会接收和整合来自多种不同感觉的输入信息，例如听觉、触觉加上视觉。注意，图1显示的皮层区域和它们之间的连接都是基于已知的研究成果，而图3则完全是个概念图，并不对应真正的皮层区域。在真正的人类大脑中，有几十个皮层区域以各种方式互相连接。实际上，大部分大脑皮层都包含联合区。这里的示意性结构图只是为了帮助你更好地了解问题本质，希望不会引起你的误解。

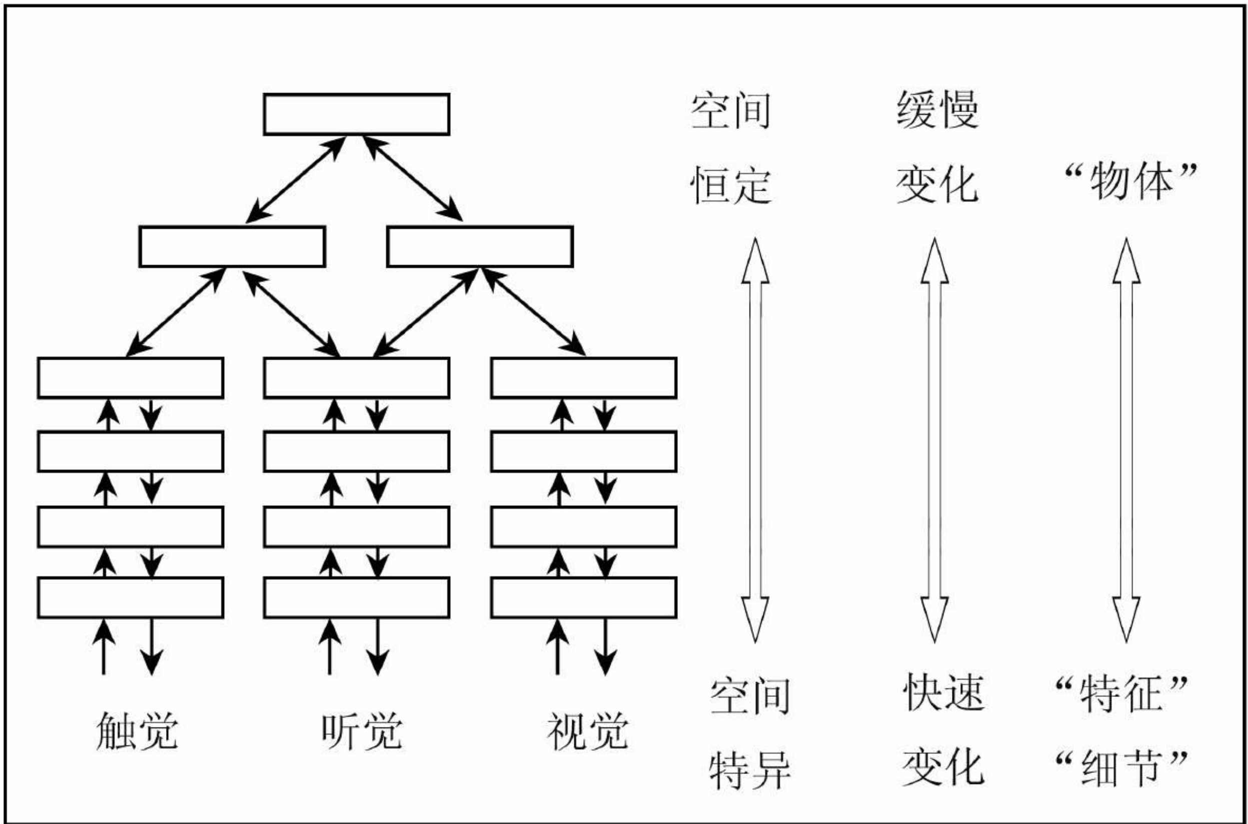


图3　触觉、听觉、视觉的恒定表征的形成

从快速变化到缓慢变化、从空间相关到空间无关，这样的转化在视

觉区域已经有大量的证据。虽然这类证据在其他感觉区相对较少，但很多神经科学家相信大脑皮层的所有感觉区都有类似的转化，而不只是在视觉区。

以听觉为例。当有人向你说话时，声压的变化非常迅速，紧接着，变化同样迅速的输入模式进入到初级听觉区A1。如果在较高层的听觉区插入探针，我们会发现一些“不变细胞”会对单词甚至短语有反应。你的听觉皮层中可能有一些细胞会在听到“谢谢你”的时候被激活，而另外一些细胞会在听到“早上好”的时候被激活。一旦你识别出这个短语，对应的细胞就会在这段说话过程中保持激活状态。

第一个听觉区接收到的模式的变化范围很广。人们可以用不同的口音、音高和速度的出同一个单词。但在更高层的听觉皮层中，低级特征不再重要，一个词就是一个词，无需考虑它的声学细节。这对音乐也是类似。《三只瞎老鼠》的曲子可以是钢琴演奏、单簧管演奏，或者是童声独唱。你的A1区接收到的模式是完全不同的。但插入在高层听觉区域的探针会发现有些细胞能一直对《三只瞎老鼠》的节奏产生反应，不管演奏乐器、节拍或其他细节有什么不同。当然，这样的实验还没有在人身上做过，因为毕竟这对被试者的侵入性太强。但是，如果你认为一定有共通的大脑皮层算法，那么你就应该能确信这种细胞的存在。我们在听觉系统中能够看到与视觉系统相同的反馈、预测和不变记忆。

最后，触觉的机制也是类似的。虽然研究者们已经开始采用高分辨率大脑成像仪，以猴子为对象进行研究，但触觉领域的决定性实验还尚未开展。当我坐在这里写作的时候，我手中拿着一支钢笔。我拿手去摸钢笔帽，手指掠过金属笔夹。当我移动手指时，虽然从我皮肤上的触觉感受器传至躯体感觉皮层的模式在快速地变化，但我仍然能感知到那是同一支钢笔。这一刻我可能在用手指掰弯金属笔夹，下一刻可能就换成另外几个手指，甚至是嘴唇。这些输入信息非常不同，会到达不同的初级躯体感觉皮层。但是，距离最初输入几层之外的较高皮层中的探针再

一次发现，有些细胞始终不变地对“钢笔”产生反应。不管我具体是用哪个手指或者身体的哪个部位触摸钢笔，这些细胞会在我触摸钢笔的过程中始终保持激活状态。

让我们想一想。对于听觉和触觉而言，我们只靠瞬间的信息输入是无法识别一个物体的。无论是来自耳朵的听觉模式还是皮肤触觉的动作模式，在任何一个时间点上都没有足够的信息告诉你在听的或接触的是什么。当你要感知一系列诸如旋律、话语或是关门声的听觉模式时，抑或当你要感知一个像钢笔那样的物体时，唯一的方式就是使用一段时间内的输入流。你不可能只听一个音调就想识别一段旋律，也不可能只碰一下就想知道那是一支钢笔。因此，感知诸如说话等事物的神经活动，一定会比单个输入模式持续的时间更长。这同之前的结论是一样的：在大脑皮层中的位置越高，随时间的变化就越少。

与听觉和触觉类似，视觉也是基于时间的输入流。但是由于我们通过一次注视就能识别物体，这就给理解带来了混乱。实际上，一次注视就能识别空间模式的能力，多年来也误导了致力于机器或动物视觉的研究者们。他们的主要问题是忽略了时间的关键性质。虽然在实验室条件下人不移动眼睛就能够识别物体，但这不是常态。在通常情况下，例如当你在阅读这本书的时候，是需要不断移动你的眼睛的。

整合感官（Integrating the Senses）

那么联合区又是怎样呢？到现在为止，我们已经看到了某个特定感觉皮层中的信息是如何上下流动的。向下的信息流动到输入端，并预测接下来会发生什么。同样的过程也发生在不同的感觉之间，也就是在视觉、听觉和触觉之间。例如，当听到一些声音后，我会预期应该看到或触摸到什么。现在，我正在卧室里写作。我家的猫Keo有个铃铛会在它走动的时候响起来。我听到她的铃铛声从走廊传来。通过这个听觉输

入，我认出了我的猫，然后我将头转向走廊，看到Keo走进来。我是通过声音来预测将要看到她的。如果Keo没有走进来，或者是其他动物出现了，我就会很吃惊。在这个例子中，听觉输入让我通过声音识别出Keo。该信息通过听觉皮层区流动到连接视觉和听觉的联合区。紧接着，相关表征会流回到听觉和视觉皮层区，进行听觉和视觉预测。图4就展示了这个过程。

这种跨感官预测很常见。我掰起钢笔金属夹，然后将手指放开，我预测会听到金属夹"啪"的一声打到笔帽上。如果没有听到这个声音，我就会很吃惊。我的大脑能够准确地预测听到声音的时间，以及听到什么样的声音。相关信息会首先通过躯体感觉区流到更高的联合层，然后流回到躯体感觉区和听觉皮层，从而预测我会听到并感觉到金属夹拍击笔帽的过程。

另外一个例子是，我每周有几天会骑自行车上班。那些天的早晨，我会去车库取车，把它转过来，推到车道上。在这个过程中，我会接收到许多视觉、触觉和听觉输入信息。自行车碰到门框，车链发出噪音，踏板碰到我的腿，车轮在地板上滚动。在我把自行车推出车库的过程中，我的大脑会不断地接收视觉、听觉和触觉信号。每一个感觉输入流都以惊人的协同方式预测其他方面的感觉。我看到的东西会帮助我预测将要碰到或听到的东西，以此类推。当我看到自行车碰到门框的时候，我会预测会听到碰撞声，并感觉到自行车被反弹起来。当我感到踏板碰到了我的腿，我会向下看，并预期会看到踏板就在我所感受到的位置。这些预测非常准确，以至于如果有任何一点不协调或反常都会被我注意到。这些信息在不同的感觉皮层区中同步地向上向下流动，创造出多感官之间统一的体验和预测。

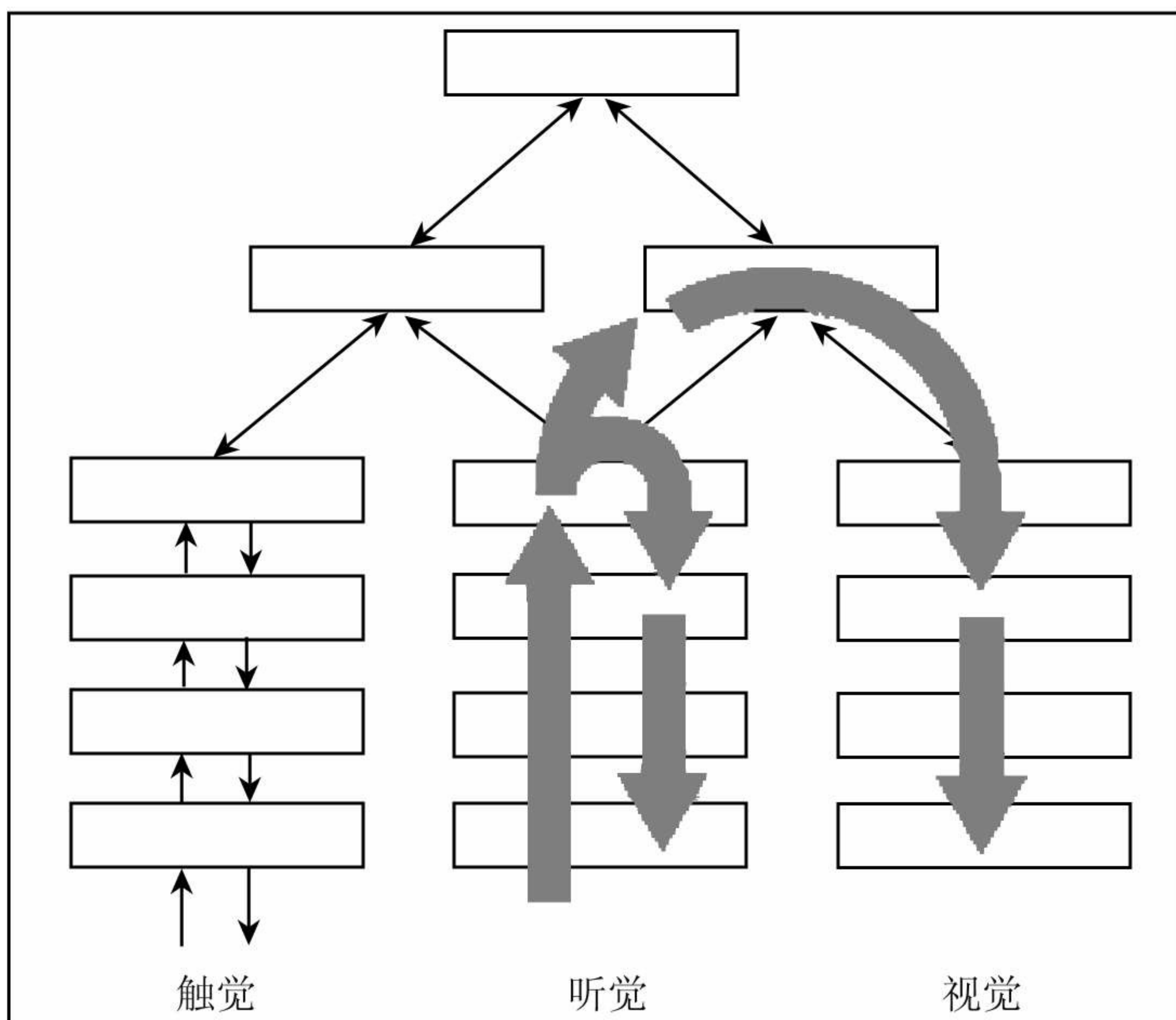


图4　信息通过感知层状结构上下流动形成预测，并创造统一的感知体验

再试试这个实验。让我们停止阅读，站起来活动活动身体、接触些东西。例如，走到水槽前面打开水龙头。在你这样做的时候，注意每个声音、触感和视觉变化。你必须很专注，你的每个动作都与视觉、听觉和触觉密切相关。当你扭动水龙头的时候，你的大脑会预期感受到皮肤上的压力和肌肉承受的阻力。你预期会看到和感受到水龙头的移动，看到和听到水流出来。当水落到水槽中的时候，你预测会听到不同的声音，看到水花飞溅。

不论是否意识到，你总会预见到自己每走一步都会发出声音。即使像拿着这本书这样简单的动作，也会涉及很多感知预测。假如你听到书

本合上的声音，但却看到它仍然打开着，你会非常震惊和困惑。正如我们在第五章中那个改变门的思想实验里所看到的那样，你会不断地对世界作出在各个感觉上协调一致的预测。当专注于这些细微感觉时，我对于这些知觉预测能够如此完全地整合在一起而感到惊叹。虽然这些预测看上去很简单甚至琐碎，但考虑到它们的无处不在并且只有基于在大脑皮层中上下流动的大量模式的协调一致下才能产生，你大概就不会这么想了。

一旦理解感觉之间的相互关联是如何的，你自然能得到如下结论，即整个大脑皮层、所有的感觉区和联合区都是一个整体。是的，你有一个视觉皮层，但它只是整个感觉系统的一部分而已。这个感觉系统是一个包含多个分支的层次结构，视觉、听觉、触觉乃至更多感觉信息都在其中上下流动。

此外还有一点，所有的预测都是通过经验学习得到的。我们现在或将来之所以预测笔夹在击打笔帽的时候会发出声音，是因为过去就是这样发生的。自行车在车库中发生碰撞的景象、触觉和声音都如我们所预期。你生来并不具备这些知识，你是利用了超大容量的大脑皮层来记住了这些模式。如果在流入大脑的输入信息中包含一致的模式，大脑皮层就会用它们来预测未来可能发生的事情。

虽然图3和图4并没有画出运动皮层，但你可以想象它们就像感觉皮层一样，是另外一个带有层次结构的“煎饼”，通过联合区与感觉系统互联（运动皮层与躯体感觉皮层的连接更紧密，这是为了便于实现躯体动作）。通过这种结构，运动皮层几乎与感觉皮层有相同的工作机制。任何一个感觉区的输入信息都能上流到联合区，形成模式后再向下流到运动区，从而产生行为。正如一个视觉输入会形成模式流到听觉或触觉区那样，它也会流到运动区。在前一种情况中，我们将这种向下流动的模式理解为预测。而在运动皮层中，我们将其理解为运动指令。正如蒙卡斯尔（Mountcastle）所指出的，运动皮层看起来就像感觉皮层一样。因

此，大脑皮层处理感官预测的方式类似于它处理运动指令的方式。

我们很快会看到，大脑皮层中并不存在单纯的感觉区或运动区。感觉模式会同步地流向任何地方，然后流回到层状结构的其他区域，做出预测或运动行为。虽然运动皮层有一些特别性质，但把它当作整个层次记忆预测系统的一部分仍然是对的。它就像另外一种感官而已。视觉、听觉、触觉以及行动深度交织在一起。

关于V1区的新观点（A New View of V1）

解开大脑皮层架构的下一步需要我们用一种新的方式来看待大脑皮层区域。我们知道大脑皮层体系的较高区域会形成恒定表征。但为什么这样重要的功能只发生在较高区域呢？受到蒙卡斯尔关于对称的观念之启发，我开始探索大脑皮层区域之间的不同连接方式。

图1展示的是视觉通路中的4个典型区域，V1区、V2区、V4区和IT区，其中V1区位于结构的最底层，然后是V2区和V4区，而IT区在最高层。传统观念认为每一层都是单独而连续的。因此，所有的V1区细胞都被认为具有相似的功能，虽然分别负责视野的不同部分。V2区的所有细胞也都执行相同的任务，V4区的细胞也互相类似。

按照这个传统观点，当人脸图像进入到V1区的时候，V1区的细胞会用线段和其他基本特征创建出这张脸的草图。这个草图被传到V2区，V2区的细胞对人脸图像进行更精细的面部特征分析后，将其传给V4，以此类推。只有当输入信息进入最高层IT区后，才会对物体建立恒定表征和进行识别。

不幸的是，这些关于较低层次的V1区、V2区和V4区的观点有一定的问题。还是那个老问题：为什么恒定表征只发生在IT区？如果大脑皮层的所有区域都有相同的功能，为什么IT就得这么特殊呢？

其次，人脸可以出现在V1区的左侧，或者右侧，你都能将它识别。但实验清楚地表明，V1区中不相邻的部分是没有直接连接的，V1区的左侧区域无法直接知道右侧区域看到了什么。让我们退一步再想想这个问题。V1区的不同部分显然在执行类似的功能，因为他们都参与了识别一张人脸。但同时它们在物理上又是相互独立的。V1区中的不同区域在物理上互相独立，但却在做相同的事情。

最后，实验表明，大脑皮层的较高区域会整合来自多个较低感觉区域的输入（如图3所示）。在真正的大脑中，一个联合区会汇聚十几个低层区域。但是在传统观念看来，像V1区、V2区、V4区这样的低层感觉区会有不同类型的连接。每个感觉区看上去只有一个输入来源，即只有自下而上的单向流动，没有明显的来自多个低层区域的输入整合。也就是说，V2区仅从V1区得到输入。为什么一些大脑皮层区域能够整合多个输入信息，而其他的则不能的？这也同蒙卡斯尔的通用大脑皮层算法的想法不一致。

由于这些原因，我开始相信V1、V2和V4不应该被看作单独的区域，而是每个都包含多个子区域。让我们回到将大脑皮层比作平铺的"餐巾"的比喻上来。如果我们要用笔在这个餐巾上标记出大脑皮层所有的功能区，目前最大的区域是初级视觉区V1，其次是V2。他们比绝大部区域都要大。我所建议的是，V1区应该被划分为很多小的区域，也就是说，不再是餐巾上的一大块区域，而是在过去分配给V1的区域中划分出很多小的区域来。换句话说，在所有视觉区域中，V1区含有最多的独立子区域。V2中的独立子区域相对较少，但面积比V1区中的略大一些。对V4区也是如此。但位于最高层的IT区却是一个完整的单独区域。这也是为什么IT区的细胞对于整个视觉世界具有鸟瞰的视野。

这里有一个令人愉快的对称现象。图5展示了如图3的层次结构，不同之处在于它展示的是我刚才描述的层次结构。注意，现在大脑皮层的每个地方看上去都是相似的。挑出任何一个区域来看，都能发现很多较

低层感觉区域为它提供输入信息。接收输入信息的区域会将反馈发送给它的输入区域，告诉它们接下来应该预测会看到什么模式。较高层的联合区会从多个感觉区（例如视觉和触觉）整合信息。而像V2子区域这样的较低层区域，会整合来自V1区中的多个独立子区域的信息。每个区域都不知道——实际上也无法知道——这些输入信息的意义。一个V2子区域不必知道它正在处理来自多个V1子区域的视觉输入信息。一个联合区不需要知道它正在处理来自视觉和听觉的输入信息。确切地说，任何一个大脑皮层区域的工作就是要弄清输入之间的相互关系，记住这些关联序列，并利用这些记忆来预测输入信息将会发生什么。大脑皮层就是大脑皮层。每个区域都进行着同样的加工过程，这就是通用大脑皮层算法。

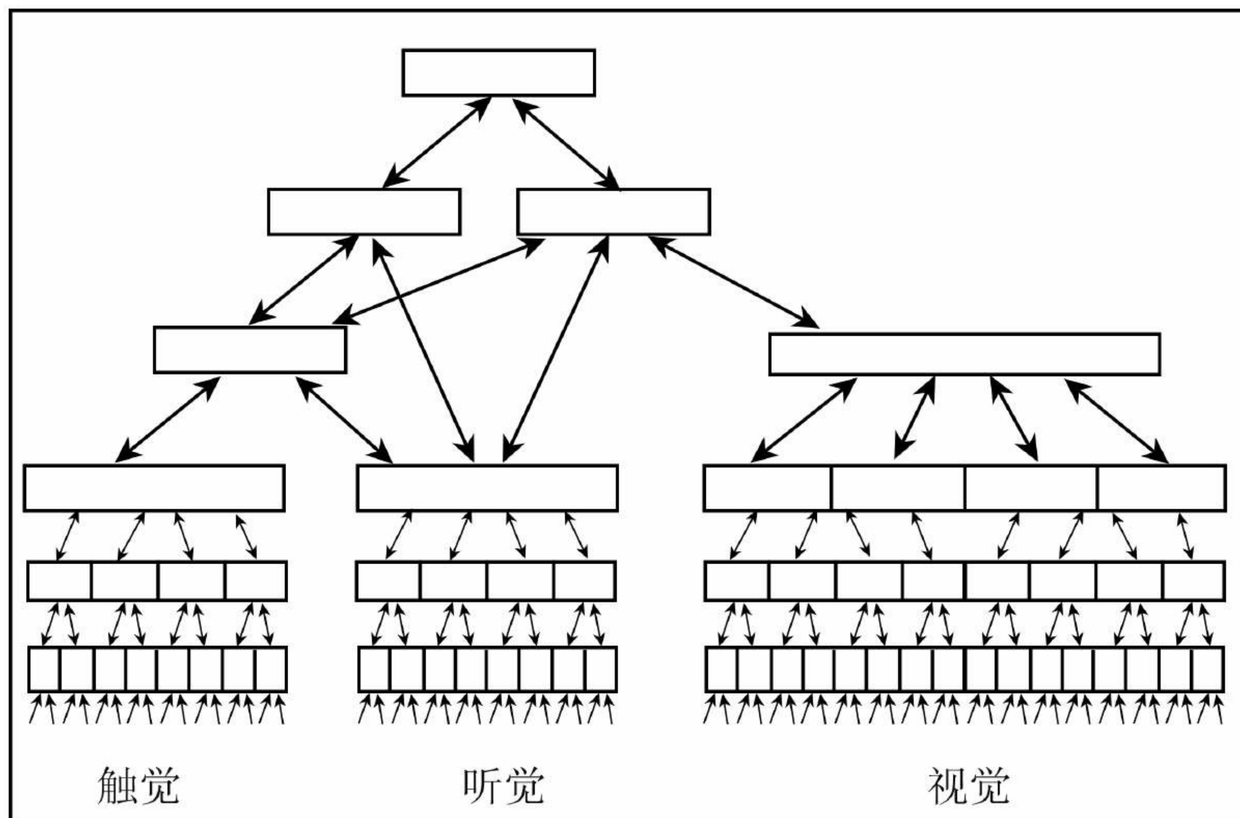


图5　大脑皮层层次结构的另外一种视角

这个新的层次结构描述能够帮助我们理解创造恒定表征的过程。让我们看看它在视觉区域中是如何工作的。在视觉处理的第一个层面上，视野的左侧和右侧是不同的，就好比听觉不同于视觉一样。V1区中的

左右两侧子区域之所以能产生相同的表征，是因为它们在现实生活中曾接触过相似的模式。就像听觉和视觉那样，它们可以被看作是独立的感官信息流，在较高层次中得到整合。

类似地，V2区和V4区中的子区域都是视觉的联合区。（子区域之间可能有重叠，但这不会从根本上影响这些子区域的工作方式。）这样理解视觉皮层与相关的解剖学知识并不冲突，也没有改变解剖学什么。信息顺着这个层级记忆树的所有分支上下流动。来自左侧视觉区域的模式会为右侧视觉区域带来预测，就像我家猫的铃铛会引发我作出“它进入了我的卧室”的视觉预测那样。

关于我们描述的这个新大脑皮层层次结构，它的最重要结果是，现在我们可以说，大脑皮层的每一个区域都形成恒定表征。在旧的观点中，只有等输入信息到达了最高层区域，也就是鸟瞰整个视觉世界的IT区时，我们才能形成诸如“人脸”的完整的恒定表征。但现在我们可以说，恒定表征是无处不在的。恒定表征存在于每个大脑皮层区域。表征的恒定性并不是到了较高的皮层区域（如IT区）才突然冒出来的。每一个区域都从位于自己下级的输入区域中提取恒定表征。因此，V4区、V2区和V1区的子区域都会基于进入它们的输入流建立恒定表征。它们可能只看到世界的一小部分，它们所处理的感官对象的“词汇”也是更为基本的，但它们执行的工作与IT区相同。此外，IT区之上的联合区同样会根据多感官输入形成恒定表征。因此，大脑皮层的所有区域都在形成层级体系中位于自己下方的那个世界的恒定表征。大脑皮层的所有区域都在形成关于这个世界的恒定表征，这些恒定表征是带有层次结构的。此中有真美。

我们的谜题发生了转移。我们不需要再问恒定表征是如何通过自底向上的四阶段形成的。但是，我们需要问，在每个单独的大脑皮层区域中恒定表征是如何形成的。假如我们认为通用大脑皮层算法是存在的，这个问题就很有意义。如果一个区域会存储模式序列，那么每个皮层都

会存储模式序列。如果一个区域会建立恒定表征，那么所有的区域都会建立恒定表征。如图5这样绘制的大脑皮层的层次结构，让这种解释成为可能。

世界的模型（A Model of the World）

为什么大脑皮层是有层级结构的？

你能够思考、运动，以及预测未来，都是因为你的大脑皮层建立了一个关于这个世界的模型。这本书最重要的一个概念就是，大脑皮层的层级结构存储了这个世界的层级结构模型。你的大脑皮层的嵌套结构反映了这个世界的嵌套结构。

嵌套或层级结构是什么意思呢？考虑一下音乐，音符组成了音程，音程组成了旋律，旋律组成了歌曲，歌曲组成了专辑。再考虑一下书面语。字母组成了音节，音节组成单词，单词组成句子。再看其他方面，想想你家周围，可能会有道路、学校和房屋，房子会有房间，每个房间都有墙壁、天花板、地板、门、和窗户。而它们每个又由更小的物体组成，例如窗户由玻璃、框架、插销和纱窗组成，而插销则由螺丝等更小的部件组成。

再花点时间看看你的周围。从视网膜进入初级视觉皮层的模式相互组合形成线段。线段相互结合形成更复杂的形状。这些复杂形状相互结合形成鼻子等。鼻子、眼睛和嘴相互组合形成人脸。人脸和身体其他部位相互结合，形成了房间中坐在你对面的那个人。

在你的世界中，所有的对象都由更小的对象组成，它们有机地组合在一起，这就是一个对象的定义。当我们给一个对象起名字，是因为它的很多特征经常出现在一起。一张人脸之所以是一张人脸，是因为两只眼睛，一个鼻子和一张嘴总是一起出现。一只眼之所以是一只眼，是因

为瞳孔、虹膜和眼睑等总是一起出现。椅子、汽车、公园和国家同样如此。最后，一首歌之所以是一首歌，是因为一系列的音程总是按次序连续出现。

因此，这个世界就像一首歌那样。这个世界的每个对象都由更小的对象组成，而大部分对象是更大对象的一部分。这就是我所谓的嵌套结构。一旦你意识到这一点，你会发现到处都是嵌套结构。而你关于世界的记忆以及大脑对它们的表征，都以完全类似的嵌套结构存储在大脑皮层的层级结构中。你关于你家的记忆并不是存在一个区域中。它存在于大脑皮层的层级结构中，而这又反映了你家的层级结构。大规模的关联关系存储在高层区域，而小规模的关联关系则存储在低层区域。

大脑皮层的设计和学习方法自然能够发现世界的层级关系。你生来并不具备关于语言、房屋和音乐的知识。大脑皮层能够智能地发现层级结构的存在，并捕捉到它。如果结构消失了，我们就陷入混淆，甚至进入一片混沌。

在任何时刻，你都只能体验这个世界的一部分。在某一时刻，你只能在你家的一个房间里，看着一个方向。由于大脑皮层的层级结构，你能够知道你是在家里，在你的客厅，看着一个窗口，即使这时你的眼睛只不过正在注视着窗户的一个“插销”而已。大脑皮层的高层区域正维持着关于你家的表征，而较低区域则维护着房间的表征，再低的则维护着窗户的表征。同样，即使在任何时间点上你都只听到一个音符，这本身无法告诉你任何信息，但层次结构可以让你知道你正在听一首歌，一张专辑。层次结构也可以让你知道你正跟你最好的朋友在一起，即使你的眼睛正在注视她的手。你大脑皮层的较高区域一直在关注全局，而较低区域则积极地处理瞬息万变的微小细节。

由于在任一时刻我们只能感到、听到和看到这个世界的很小一部分，流入大脑的信息自然而然地是以模式的序列到达的。大脑皮层希望学习到那些反复出现的序列。某些情形下模式序列有比较严格的次序，

例如音乐旋律的音程顺序。我们大多会对这种序列很熟悉。但我这里所说的模式序列包含更宽泛的情形，更像是个数学集合。序列是一个包含多个模式的集合，模式互相伴随出现，次序经常并不固定。与出现次序相比，更重要的是这些模式会互相紧密地跟随出现。

要讲清楚这个问题我们需要一些例子。当我看你的脸的时候，我接收的输入模式序列并不是固定不变的，而是由我的扫视决定。这次我扫视的顺序可能是“眼睛——眼睛——鼻子——嘴”，下次可能是“嘴——眼睛——鼻子——眼睛”。这些脸部器官构成一个序列。它们在统计上相互关联并经常一起出现，虽然出现次序可能会不同。如果你感知到一张人脸，而此时你正注视着“鼻子”，那么接下来的模式很有可能是“眼睛”或者“嘴”，而不是“钢笔”或“汽车”什么的。

大脑皮层的每个区域都会看到这种模式流。如果区域能够学习到模式之间的这种关联关系，并能够预测接下来会发生什么，那么大脑皮层区域就形成了该模式序列的持久表征或者记忆。学习序列是形成对真实世界对象恒定表征的最基本的要素。

真实世界的对象可以是具体的，如蜥蜴，人脸或一扇门，也可以是抽象的，如一个字或一个理论。大脑以相同的方式处理抽象对象和具体对象。它们都不过是以可预见的方式出现在一起的模式序列。正是这些反复出现的输入模式，让大脑皮层区域知道了它们来自真实世界的某个对象。

真实的重要特征就是可预测性。如果一个大脑皮层区域发现，它能够通过一系列物理运动准确可靠地获取输入模式序列（例如用眼睛扫视或者用手指抚摸），并能够在接收到这些模式后及时准确地预测它们（例如歌曲或词语的声音），大脑就会认为它们存在因果关系。输入模式反复以相同的关系出现，要说这纯属巧合是几乎不可能的。可预测的模式序列一定是真实存在的更大对象的一部分。因此，要想了解这个世界上的几个不同事物是否互相关联，最可靠的方法就是利用可预测性。

每张脸都有眼睛、耳朵、嘴和鼻子。如果大脑看到一只眼睛，然后看到另外一只眼睛，再看到一张嘴，那么它就会很确定这是一张脸。

如果大脑皮层可以说话，它们一定会说："我处理过这么多不同的模式。有时候我无法预测接下来会是什么模式。但这些模式的确互相关联。它们经常一起出现，我能够可靠地在它们之间跳动。因此，无论我看到它们中的任何一个，都会用一个共同的名字来称呼它们。我传给更高皮层区域的就是这个组合的名字，而不是独立的模式。"

序列的序列（Sequences of Sequences）

随着信息从初级感觉区传送到更高层的区域，我们看到的变化越来越少。伴随着视网膜接收模式每秒几次的变化，V1初级视觉区中的活跃细胞集合发生着快速变化。而在IT视觉区，细胞的激发模式则非常稳定。这其中发生了什么？每个大脑皮层区域都有它熟知序列的清单，就像歌曲清单那样。区域会存储所有事物的序列：海浪拍击沙滩的声音，你母亲的脸，从你家到街角商店的路线，如何拼写单词"爆米花"，如何洗牌，等等。

每首歌都有名字，类似地，每个皮层区域知道的序列都有个名字。这个"名字"就是一组被激发的皮层细胞，这些细胞代表了这个序列中出现的对象。（现在不必关心是如何选出这组细胞并表示该序列的，我们稍后讨论。）只要序列还在，这些细胞就会保持活跃，正是这个"名字"被传给更高层区域。只要这个输入模式是可预测的序列的一部分，该区域就会将固定的"名字"传给更高层区域。

这就像这个区域在说："这是我听到、看到和触摸到的那个序列的名字，你不需要了解各个音符、边缘或纹理。如果有新的或无法预测的事情发生，我会告诉你的。"更具体地，我们可以想象视觉区最高层的

IT区向它之上的联合区报告说："我看到了一张脸。是的，眼睛的每次扫视都停留在脸的不同部位，我接连地看到这些不同部位。但这仍是同一张脸。当我看到别的东西我会告诉你的。"通过这种方式，一个可预测的事件序列会得到确定的"名字"，即某个不变的细胞激活模式。在信息沿着层次结构向上流动的过程中，这种情况反复发生。一个区域可能会识别出一个构成音素（组成单词的声音）的声音序列，并将代表这个音素的模式传给更高层区域。更高层区域识别出音素序列组成单词。再高一层的区域识别出单词序列组成短语，以此类推。需要记住的是，在最底层区域中的序列可能相当简单，例如在空间中移动的某个视觉边缘。

通过在层次结构中不断将可预测序列转变成"命名对象"，越高层级的稳定性也变得越强，从而形成恒定表征。

在模式沿着层次结构向下流动时，会发生相反的效果：稳定模式"展开"成序列。假设你七年级的时候记住了葛底斯堡演讲，现在你想背诵出来。在大脑皮层的高层语言区，会存储着一个表征林肯这一著名演讲的模式。首先，该模式会被展开成关于短语序列的记忆。在下一个低层区域，每个短语又被展开成关于单词序列的记忆。这时，展开的模式会被分别发送给听觉和运动皮层。沿着运动皮层，每一个单词又被展开成关于音素序列的记忆。在最后的低层区域，每个音素节展开成肌肉指令序列，发出声音。在层次结构中越向下看，模式的变化越快。你在运动皮层最高层看到的那个单独不变的模式，最终成为了复杂而漫长的声音序列。

在信息顺着层级结构向下流动的过程中，恒定性也具有显而易见的优势。如果你想将葛底斯堡演说用键盘打出来，而不是背诵的话，在大脑皮层最顶层中都是以同样的模式开始的。该模式在下一层被展开为短语序列。更下一层短语被展开成单词序列。到目前为止，背诵和打出葛底斯堡演说之间还没有什么区别。但在更下一层的运动皮层中，它们开

始进入不同路径。为了打字，单词展开成字母，字母展开成手指打字的肌肉指令。“87年以前，我们的祖先在这大陆上建立了一个国家……”对于这些话语的记忆被处理为恒定表征，这与你是朗诵、打字还是手写都没关系。

请注意，你不必为朗诵和打字而将这篇演讲记忆两次。关于这份演讲的记忆可以形成多种行为。在任何区域的恒定表征模式都可以沿着不同分支路径进入下层结构。

在效率方面补充一点，在低层结构中表示简单对象可以在表示不同高层序列的时候重用。例如，由于两个演讲会用到一些相同的词，我们不需要分别为葛底斯堡演讲和马丁·路德·金的“我有一个梦想”各学一组完全不同的单词。嵌套序列的层次结构允许共享和重复利用低层对象，例如单词、音节和字母。这对存储关于世界的信息和结构非常有效，它与计算机的工作方式非常不同。

与运动区类似，感觉区也同样会进行序列展开。这个过程能让你在不同视角感受和理解对象。假设你正走向冰箱去取冰激凌，你的视觉皮层在多个层次上都是激活的。在较高层，你在持续感知“冰箱”。而在较低层，这个视觉预期会被分解为若干个局部视觉输入。看见冰箱这一行为是由你对门把手、制冰器、门上的磁条以及孩子的涂鸦等的一一注视组成的。

在你的眼睛从冰箱的某个特征转到另一个特征上的几毫秒的时间里，对于扫视结果的预测也在顺着视觉皮层向下传播。只要该预测被一次次的扫视结果验证，你的高层视觉区域就会一直认为你所看到的是个冰箱。需要注意的是，与葛底斯堡演讲不同，演讲中的单词顺序是固定的，而你看冰箱的序列顺序却不是固定的。输入流和提取到的记忆模式取决于你的扫视动作。因此在这种情况下，展开模式的序列并没有严格的顺序关系，但最终结果却是相同的：缓慢变化的高层模式展开成快速变化的低层模式。

信息在大脑皮层体系上下流动的过程中，记忆和用名字表征序列的方式可能会让你联想起军事指挥的等级结构。最高将领说：“部队转移到佛罗里达州过冬。”这个简单的高层指令在沿着等级结构向下传达的过程中逐步展开成更详细的指令。将军的部下会将命令分解为若干步骤，例如准备离开、向佛罗里达州转移，以及准备到达。而其中每个步骤又会进一步分解，让下级执行。在最底层，成千上万的士兵执行成千上万的行动指令，最终完成了部队转移。每一层的执行情况都会形成报告，汇报给上级。在向上逐层汇报的过程中，报告不断汇总精简，直到最高层将领收到的每日简报称：“向佛罗里达州转移行动一切顺利。”将军不会得到行动的任何细节。

但这一规则存在一种例外情况。如果出了什么差错导致下属无法完成指令，那么这个问题会被逐层上报，直到有人知道该怎么处理。而这个知道如何处理的军官并不将其视为意外。他的部下们无法处理的意外，在他看来恰恰在意料之中，然后这个军官会给下属发出新的指令。大脑皮层有着相似的工作方式。我们将看到，当发生的事件（或者模式）并不是我们所预期的，相关信息会不断向上传递，直到有区域能够处理。如果大脑皮层的低层区域无法预测将要看到的模式，它们会视之为一个错误并将其上传，直到某个能预测到这个模式的区域。

* * *

按照设计，每个大脑皮层区域都会存储和记忆序列。但这种对大脑的描述过于简单。我们需要为模型增加一些复杂度。

大脑皮层中自下而上的输入，实际上是由上百万的神经轴突所携带的输入模式。这些轴突来自多个不同的区域，包含各种模式。即使只有1千个轴突，它们所能携带的模式类型数量也远多于这个宇宙中的分子数量。在人的一生中，一个皮层区域只能看到这些模式中很小的一部分。

所以问题来了：当一个区域存储序列时，这些序列到底是什么呢？答案是：皮层区域会首先将它的输入模式分类到若干种可能之一，然后再寻找序列。想象你是一个皮层区域。你的任务是将彩纸分类。提供给你了10个桶，每个桶上都标记了颜色，例如一个桶标了绿色，一个标了黄色，一个标了红色，等等。接下来，把彩纸一张接一张地发给你，要求你按照颜色将它们分类。你收到的纸每张颜色都略有不同。由于世界上有无限多种颜色，任何两张纸的颜色都不会完全相同。有时候，你会很容易知道这张彩纸该放到哪个桶里，有的时候却很难。一张介于红色和黄色之间的纸，既可以放在红桶里也可以放在黄桶里，但你必须要在红桶和黄桶之间选一个，即使最后选择可能是随机的。（这个例子的目的是要说明大脑必须对模式做分类，但并不是真的将模式放在桶里。）

现在你可以开始寻找序列了。你看到“红——红——绿——紫——黄——绿”经常按顺序出现，并称之为“红红绿紫黄绿”序列。需要注意的是，如果不对每个模式作分类，就不可能识别序列。不事先将彩纸分为10类，你就没法识别两个彩纸序列是否相同。

这样你就可以启动运行了。你将会看到所有的输入模式——通过大脑的低层区域传入的彩纸——对它们分类后寻找序列。分类和形成序列是产生恒定表征的两个必须步骤，大脑皮层每个区域都会这样做。

当输入有歧义的时候，形成序列的过程就显示出它的用处了，就像是将彩纸放红桶还是黄桶的问题，即使你不确定这张纸到底是红色多些还是黄色多些，你都要做这个决定。如果你了解该输入的最相似序列的话，你就可以利用这个知识帮助你作分类决定。如果你刚接收到两个红色、一个绿色和一个紫色，从而相信正在接收一个“红红绿紫黄绿”的序列，那么你就会预期下一张彩纸是黄色的。不过，但你实际接收的彩纸不是黄色，而是一种介于红色和黄色之间的颜色，甚至可能还更偏红色一些。但是你对“红红绿紫黄绿”序列更加熟悉和期待，因此你会把这张纸放入黄桶中。你这就是在用已知序列的上下文知识来消除歧义。

这种现象在我们的生活中每时每刻都在发生。当人们说话的时候，其中的很多单词如果脱离了语境就无法理解。但是当你在一句话中听到一个有歧义的单词时，你不会因为这个单词歧义而影响理解。你能够理解它的意思。类似地，手写的单词如果脱离了语境也会经常认不出，但如果是出现在一个句子中的话，就很容易被认出。大部分时候，你下意识地利用记忆填补了歧义和不完整信息。你听到你期望听到的，看到你期望看到的——至少当你看到或听到的与过去经验一致时是这样的。

注意，记忆不仅能让你消除当前输入中的歧义，还能够预测接下来会发生什么。在你的大脑皮层处理彩纸输入的同时，你还可以对负责传纸给你的“人”说：“嗨，如果你没法决定接下来该传给我什么颜色的纸，根据我的记忆，你应该把黄色的纸传给我。”通过识别模式序列，大脑皮层区域将会预测接下来的输入模式，并告诉低层区域。

大脑皮层区域不仅学习熟悉的序列，还学习如何调整分类。假如你刚开始用来分类的桶的标记分别是“绿色”、“黄色”、“红色”、“紫色”和“黄色”。你准备识别诸如“红红绿紫黄绿”之类的序列，如果其中某个颜色发生偏离怎么办？如果每次你看到的“红红绿紫黄绿”序列中，“紫色”总是奇怪怎么办？这种新颜色很可能是靛蓝色。所以你将紫桶变成“靛蓝”桶，这样，就更符合你所看到的序列了，而且降低了歧义。大脑皮层很灵活。

在大脑皮层的各区域中，自下而上的分类和自上而下的序列不断交互和变化，贯穿始终。这是学习的本质。实际上，所有的大脑皮层区域都有很强的可塑性，它们会根据经验不断改进。你记住世界的方式就是不断形成新的分类和新的序列。

最后，让我们看看这些分类和预测是如何与更高层区域交互的吧。大脑皮层的另外一部分工作是将你所看到的模式序列的名字传给更高层。也就是说你会将一张写有“rrgpog”字母的纸传给上一层。对于更高层，这些字母本身没有什么意义；它应该被看作是一个模式，与其他输

入相结合、分类，然后形成更高阶序列。像你一样，较高层也在监测所看到的序列。某个时刻，较高层可能会对你说：“嗨，如果你没法决定接下来要传给我什么，根据记忆我推测应该是‘黄黄红绿黄’序列。”这基本上告诉了你需要在你收到的输入中寻找什么。你需要尽可能地将所看到的构成这个序列。

很多人听过人工智能和机器视觉领域所使用的术语“模式分类”，让我们看看人们通常理解的模式分类过程与大脑皮层实际处理的过程有什么不同。为了让机器识别对象，研究者一般会创建一个模板——例如一个杯子，或者杯子的某种原型——然后他们会让机器将输入与这个原型杯子进行匹配。如果机器发现匹配度很高，它会说这是个杯子。但我们的大脑并不存在这种模板，每个大脑皮层区域接收的输入模式也不是图像。你并不会记得视网膜、耳蜗和皮肤所感受到的信息。大脑皮层的层次结构保证了对象的记忆是按层级分布的，而不是单独存在某个点上。而且，由于每个区域都会形成不变记忆，因此每个区域都会学习序列的恒定表征，这本身构成了不变记忆。你在大脑中不会找到关于杯子或者任何其他对象的图像。

与相机内存不同，你的大脑记住的是这个世界的本质而不是表象。当你思考这个世界，你记起来的是和对象的存在方式与行为方式有关的模式序列，而不是它们在特定场景或特定时间出现时的样子。你感受到的对象序列反映了这个世界的不变结构。你感受到的这个世界的不同部分之间的顺序是由这个世界的结构决定的。例如，你可以直接通过登机通道走上飞机，但不能通过售票台走上去。你借以感受这个世界的序列就是这个世界的真实结构，这也是大脑皮层所要记住的。

当然，不要忘记，任何大脑皮层区域的恒定表征都可以通过向下层传播而成为对感觉的细致预测。类似地，运动皮层的恒定表征也可以通过向下层传播，变成特定场景下的详细运动指令。

大脑皮层区域的结构（What a Region of Cortex Looks Like）

现在让我们将注意力集中到大脑皮层的某个区域上，也就是图5中的某个小方格。图6展示了这个区域的细节。我的目标是告诉你区域中的细胞是如何学习和回忆模式序列的，这是形成恒定表征和进行预测的关键。我们先介绍皮层区域的结构和组合方式。大脑皮层区域的大小差别很大，其中最大的是初级视觉区V1，根据它在大脑后方占据的位置，大概有护照大小。但正如前面所说，V1实际上包含了很多小的区域，每个大约只有这页纸上的一个字那么大。

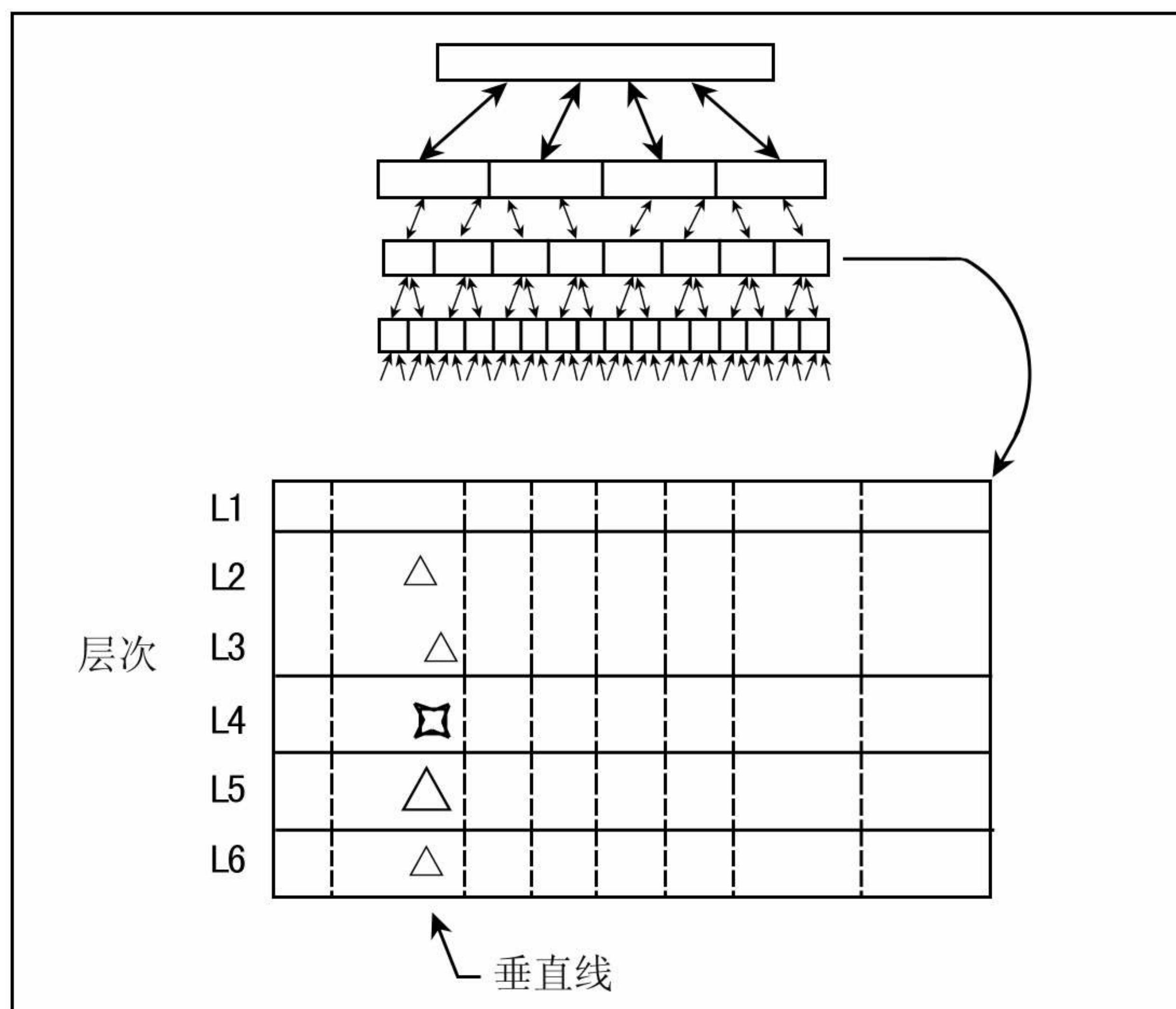


图6　大脑皮层区域中的行与列

那么现在让我们假定一个典型的区域如硬币一般大小。回忆一下我在第三章提过的6张名片，每张卡代表一层皮层组织。为什么我们认为大脑皮层存在分层呢？如果你把硬币大小的皮层区域放在显微镜下，你会看到细胞的密度和形状从上到下是有差异的。这种差异造成了分层。最顶部的第1层是6层中最独特的。它包含的细胞很少，主要由一层平行于皮层表面的神经轴突组成。第2、3层比较类似，主要由很多紧挨在一起的金字塔形细胞组成。第4层由星形细胞组成。第5层既有一般的金字塔形细胞，还有一种特别大的金字塔形细胞。最下面的第6层也有几种独特的神经元细胞。

我们可以看到的是这些水平的分层结构，但很多时候科学家们还会经常提及垂直于分层结构的细胞垂直柱结构。你可以将这些垂直柱想象成一起协同工作的纵向细胞单元。（"垂直柱"这一术语在神经科学领域存在很多争议。对于它们的大小、功能和重要性都有争论。对我们而言，你只需要将其大致想象为某种柱状结构就可以了，这一点大家都是认可的。）每个垂直柱中的不同分层都通过上下延伸的轴突互相连接，并形成神经突触。垂直柱并不真的像一个个柱子那样有明显分界——大脑皮层不会这么简单——但垂直柱的存在可以通过几个现象推断出来。

第一个证据是每个垂直柱中的细胞会倾向于对同一刺激产生激活。如果我们仔细看V1区中的垂直柱，我们会发现其中一些垂直柱会对朝某方向倾斜的线段（/）发生反应，而另一些会对朝另一个方向倾斜的线段（\）发生反应。每个垂直柱中的细胞都紧密互联，这也是为什么它们整体会对相同刺激产生反应的原因。更具体来讲，在第4层的激活细胞会让在它之上的第3、2层的细胞激活，然后又会让在它之下的第5、6层的细胞激活。信息在同一个垂直柱的细胞中上下传播。

另一个证据是大脑皮层的形成方式。在胚胎期，单一的前体细胞（precursor　cell）会从内脑腔移动到皮层最终形成的地方。每个这样的细胞会分裂成为大约100个神经元细胞，我们称之为微型柱

（microcolumn），这些细胞像我刚才所介绍的那样纵向互联。“垂直柱”这个词经常被用来描述不同的现象，它既可以指纵向连接，也可以指发育自相同源细胞的一组细胞。如果采用后一种定义，人类大脑皮层大约有几亿个微型柱。

为了帮助你对这种柱状结构产生直观认识，你可以想象单个微型柱的粗细同我们的头发一样。取几千根头发，将它们切成小段——就像去掉点的小写字母i那么长。我们把这些头发并排粘起来，看起来像一个很稠密的刷子。然后再将非常细的长头发粘合成一层——用来表示你第1层的神经轴突——然后将这层头发水平地粘在刚才的短发上。这个像刷子样的东西就是你那硬币大小的皮层区域的简化模型。信息基本上沿着头发的方向流动：在第1层沿着水平方向，而在第2到第6层沿着垂直方向。

关于垂直柱你还需要了解另一个细节，然后我们才能来谈这种结构的作用。靠近观察，你会发现每个垂直柱中的细胞上90%的突触是与外部相连的。其中一些突触来自相邻的垂直柱，另外一些则来自半个大脑以外的位置。那么，既然有这么多的皮层连接广泛分布在水平方向，我们为什么要强调垂直柱的重要性呢？

答案就在这个记忆-预测模型中。在1979年，当弗农·蒙卡斯尔提出通用大脑皮层算法时，他还提出大脑皮层的垂直柱是皮层计算的基本单元。但是，他并不知道垂直柱的功能是什么。我认为垂直柱是进行预测的基本单元。如果一个垂直柱想要预测何时该激活，它就需要知道大脑其他地方的情况，因此才会有来自各个地方的神经突触连接。

我们接下来将要涉及更多细节，这里可以先说明一下大脑为什么需要这种连接。为了预测一首歌的下一个音符，你需要知道这首歌的名字，现在播放到了哪里，从上个音符至今已经过了多久，以及上个音符是什么。每个垂直柱中的大量突触，可以将这个垂直柱中的细胞与大脑的其他部分联系起来，从而为这个垂直柱提供它需要的上下文信息，来

预测不同情况下的行为。

* * *

接下来我们需要考虑的是，这些硬币大小的区域（以及它们的垂直柱）是如何通过层级结构上下接收和发送信息的。我们首先看上行部分，如图7所示，这是一个相对比较直接的流动过程。假设我们正在看一个皮层区域以及它的几千个垂直柱。我们放大只看其中一个垂直柱。从底层区域收集的输入总是会到达第4层，即主输入层。在传播过程中，这些输入信息还会在第6层形成连接（稍后我们会看到这为什么很重要）。第4层的细胞会将这些输入信息投射到第2、第3层。当一个垂直柱向上层投射信息的时候，第2、第3层的细胞会通过轴突传给更高层区域的输入层。信息就是这样沿着皮层区域向上传递的。

如图8所示，下行信息流的路径相对不那么直接。皮层垂直柱中的第6层细胞是向下投射的输出细胞，它们将投射信息投射到下层区域中的第1层细胞。而在下层区域的第1层中，细胞轴突又会延伸很长的距离。因此，从一个垂直柱下行的信息会有可能到达并刺激多个较低层区域的垂直柱。在第1层只有很少的细胞，但是第2、第3、第5层的细胞都在第1层有树突，因此这些细胞会被流经第1层的反馈所激活。来自第2、第3层细胞的轴突在离开大脑皮层的时候会在第5层形成突触，会激活第5、第6层的细胞。因此，我们可以说信息下行的路径并没有那么直接。它会在第1层扩散的时候进入多个不同方向。反馈信息在较高层区域的第6层细胞形成，然后通过较低层区域的第1层扩散。在较低层区域的第2、第3、第5层的一些细胞会被激活，其中部分又会激活第6层的细胞，并进一步投射到更低层区域的第1层，如此反复。（如果你仔细看图8，这个过程会更容易理解。）

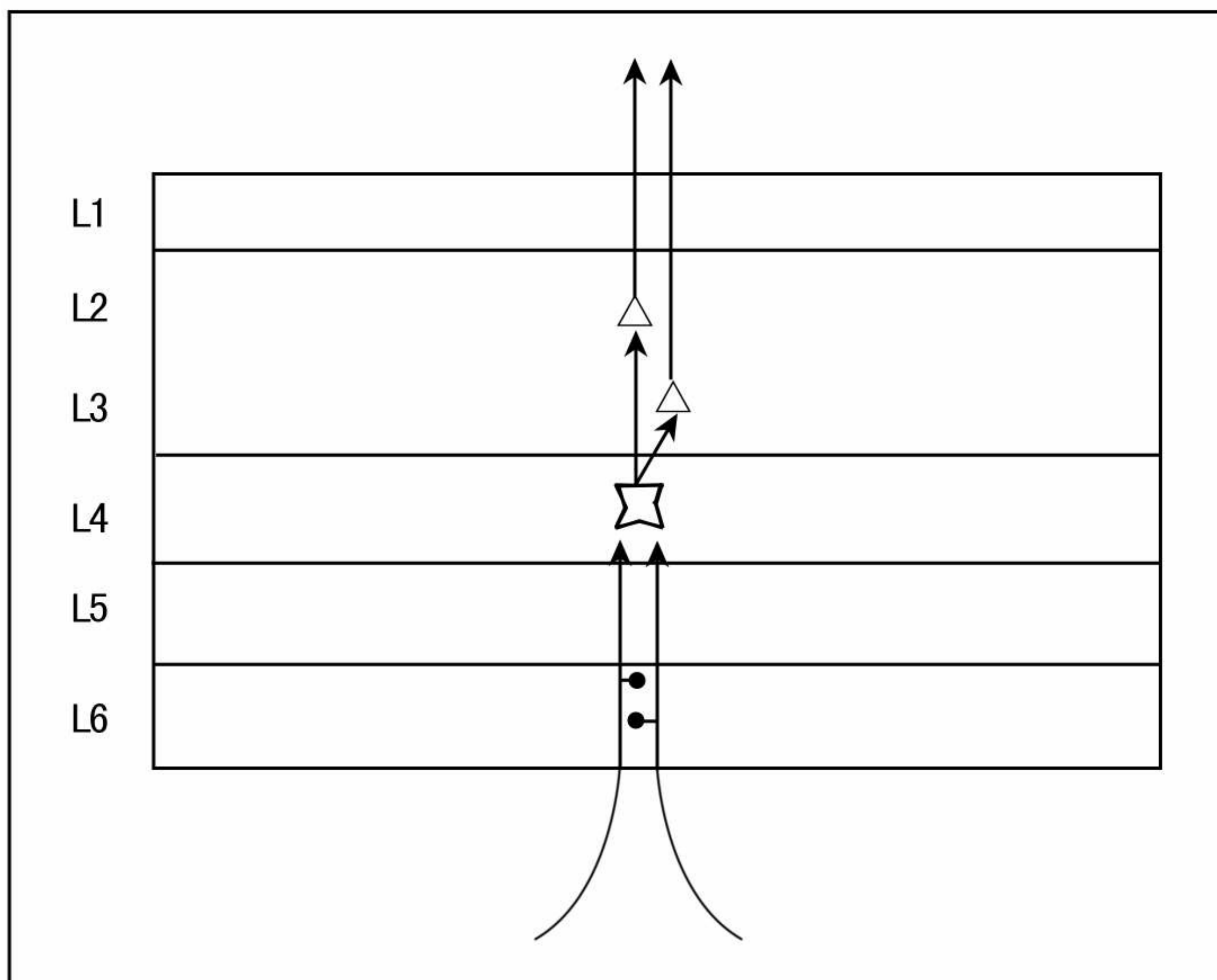


图7　大脑皮层区域中信息的上行路径

这里我们先看一下信息为什么会在第1层扩散。要将恒定表征转换为特定预测，需要有随时决定信息流沿哪条路径向下传播的能力。第1层提供了一种将恒定表征转变为更详细的特定表征的方式。我们前面提到，无论用口语还是书面的方式，你都能回忆起葛底斯堡演讲。这意味着，同一个通用表征可以沿着两条路径移动，一个是面向口语，而另一个面向书面。类似地，当我听到旋律中的下一个音符时，我的大脑会选取一个通用音程，如五度音程，然后将其转换为正确的特定音符，如C调或者G调。第1层中水平流向的神经活动就提供了这种机制。高级别的恒定预测若要沿皮层向下传播成为特定预测，必须要有一种机制能够保证，每个层级上的模式都能流入不同分支。第1层满足了这个要求。即便我们不知道第1层的存在，也依然能够预测出这一机制的必要性。

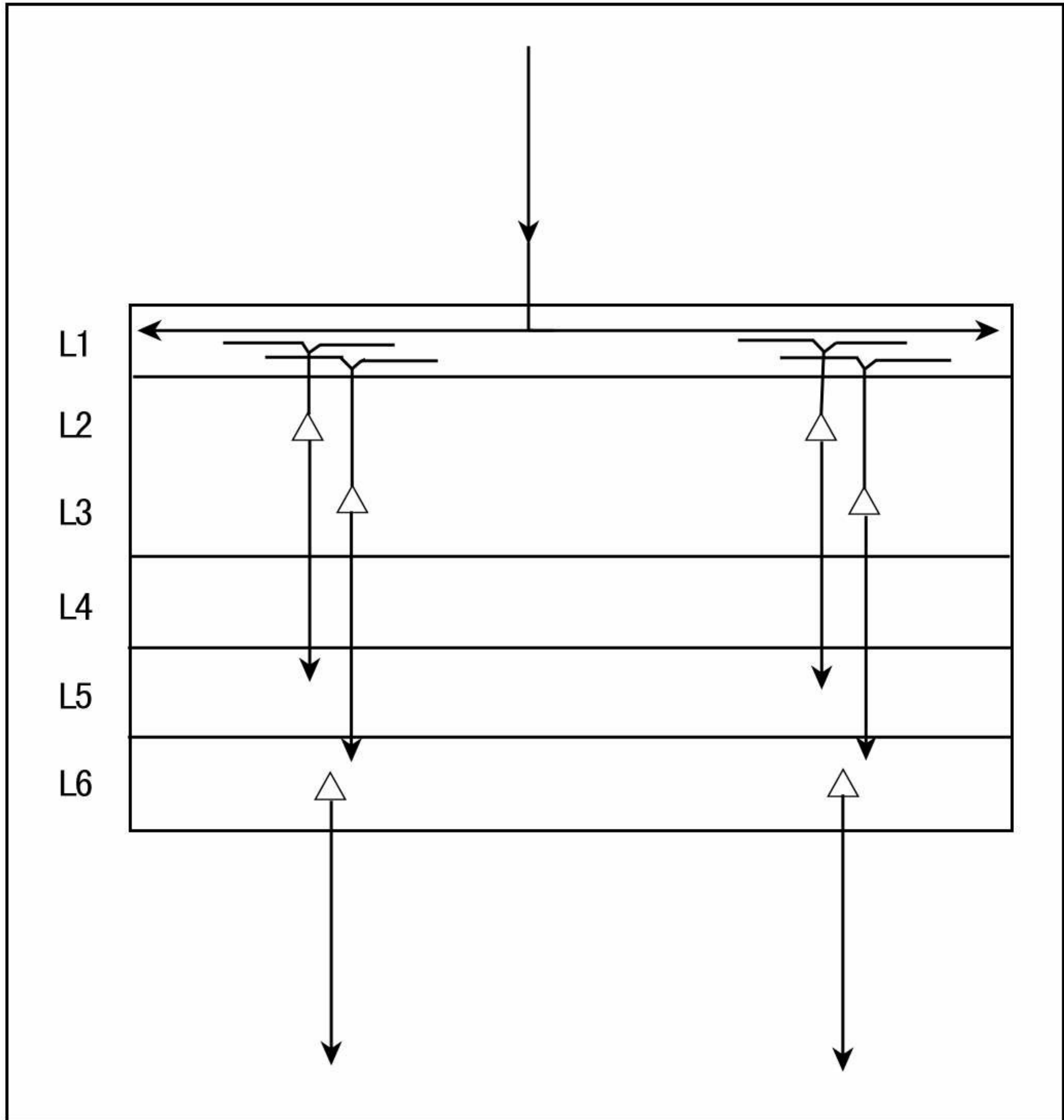


图8　大脑皮层区域中信息的下行路径

最后介绍一点解剖学知识：当轴突离开第6层到达其他地方的时候，它们被包裹在一层名为髓磷脂的白色脂肪物质里。这种脑白质就像是你家里电线的绝缘层。它有助于防止信号互相干扰，并能让它们传播得更快，时速可达每小时两百英里（约320公里）。当轴突离开白质时，它们就进入了新的皮层垂直柱的第6层。

* * *

最后，大脑皮层区域之间还有一种间接方法来产生交互。

在我介绍该细节之前，请回忆一下关于第二章探讨过的“自-联想记忆”。你应该还记得，自联想记忆可以用来存储模式序列。当一组人造神经元的输出信息被重新作为输入信息反馈给所有神经元，而且还增加了一个延时，那么这些模式就会一个接一个地形成序列。我认为大脑皮层也是采用相同机制来存储序列的，虽然会有一些额外变动。我们不是用人造神经元来形成自联想记忆，而是利用皮层垂直柱。所有垂直柱的输出都重新送入第1层。这样，第1层就包含了以下信息：该皮层区域中哪些垂直柱刚才是激活的。

如图9所示，让我们来考察一下其中的元素。很多年前我们就已经知道，运动皮层（M1）中的第5层大细胞与你的肌肉以及脊髓中的运动分区存在着直接的联系。这些细胞能驱动你的肌肉让你运动。无论你是在说话、打字还是做任何复杂动作，这些细胞都会高度协同地不断激活和抑制，让你的肌肉收缩。

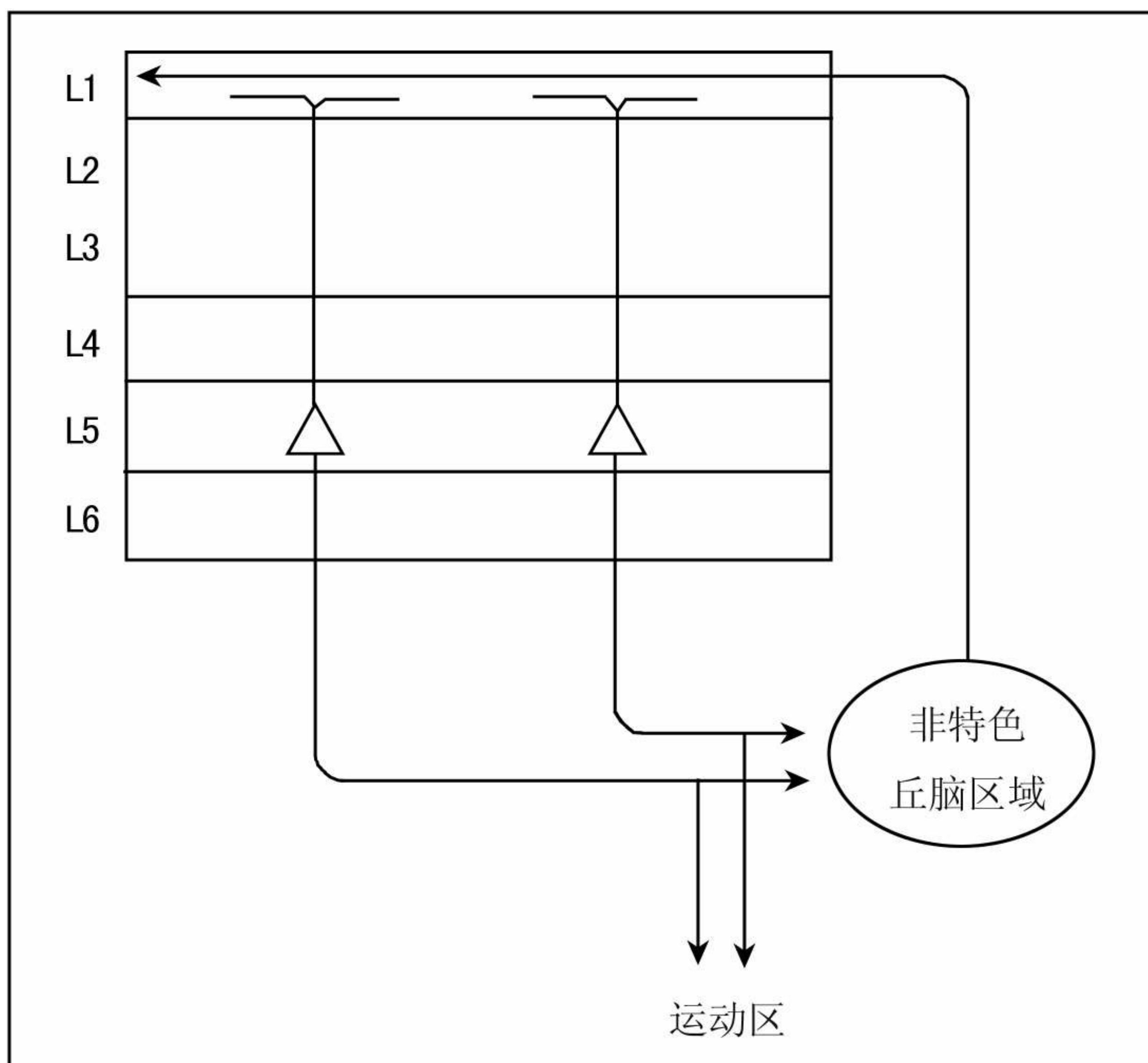


图9　当前状态和运动是如何通过丘脑通信的

着重要角色，而不仅仅是在运动区域。例如，视觉皮层的第5层大细胞会投射到控制眼球移动的那部分脑区。所以，视觉皮层区域，如V2区和V4区，不仅负责处理视觉输入，还会协助控制眼球的移动，进而决定你能看到什么。在大脑皮层的每个区域中都遍布第5层大细胞，这表明它们在各种类型的运动中都发挥作用。

除了在运动方面发挥重要作用，第5层大细胞的轴突还分为两个分支。其中一个分支进入大脑的丘脑区，如图9中右下角的圆圈所示。人

类的丘脑长得像两颗鸟蛋，大小也差不多。它位于大脑的正中央，在旧脑之上，被白质和大脑皮层包围着。丘脑接收来自大脑皮层每个区域的轴突，同时它也有轴突延伸回那些区域。这些连接的很多细节已经很清楚了，然而丘脑仍然是一个复杂结构，我们还不清楚它的作用。但是丘脑对人们的正常生活很重要，如果丘脑受伤会让人进入永久的植物人状态。

从丘脑到大脑皮层的路径有很多，但现在我们只对其中的一条感兴趣。这条路径起始于第5层大细胞，它们连接到丘脑的某类非特定细胞上。这些非特定细胞又会将轴突投射回大脑皮层不同区域的第1层中。例如，V2和V4区域的第5层细胞会将轴突送到丘脑，然后丘脑会将信息发回到V2、V4区域的第1层。大脑皮层的其他部分也是这样。皮层中多个区域的第5层细胞将信息投射到丘脑，然后丘脑会将这些信息反馈给这些区域和相关区域的第1层中。我认为这个回路正像是自－联想记忆中能够学会形成序列的延时反馈。

到目前为止，我已经谈及了第1层的两个输入模式。皮层中的较高层区域会将兴奋扩散给较低区域的第1层。而同一区域中的兴奋垂直柱同样会将兴奋通过丘脑传送给第1层。我们可以将第1层收到的这些输入信息看作歌曲名称（来自较高区域的输入信息）以及我们正在听歌曲中的什么位置（来自同一区域的兴奋垂直柱的延时反馈）。因此，第1层中承载着能让我们预测一个垂直柱何时应该兴奋的大量信息——包括序列的名字以及我们在序列中的位置。利用第1层的这两种信号，一个皮层区域就能够学习和回忆模式序列了。

大脑皮层区域如何工作：相关细节

有了这3种回路——沿皮层体系向上的模式会聚，沿皮层体系向下的模式发散，以及通过丘脑形成延时反馈——我们就可以来看一个皮层

区域是如何完成皮层所需的功能了。我们想知道的是：

1．大脑皮层区域是如何将输入模式分类的（就像彩色纸片桶那样）？

2．它是如何学习模式序列的（例如旋律的音程或者人脸的“眼睛——鼻子——眼睛”的序列）？

3．它如何形成一个序列的恒定模式或者“名字”？

4．它如何作出特定预测的（在恰当的时间迎接火车，或者预测旋律中的特定音符）？

让我们首先假设皮层区域中的垂直柱就像我们分类彩纸时所用到的桶。每个垂直柱代表一个桶上的标签。每个垂直柱中的第4层细胞从几个较低层区域接收输入信息，如果这些输入信息的组合是正确的，这个垂直柱就会被激活。当第4层细胞激活，就表明它认为输入信息与其垂直柱的标签相符。正如彩纸分类那样，输入信息可能是有歧义的，因此可能会有多个垂直柱都能匹配上这个输入信息。我们希望皮层区域能消除歧义，判断彩纸到底属于红色还是黄色，而不是两者兼有。一个激活程度较强的垂直柱应当能够抑制其他垂直柱的激活。

大脑中有抑制性细胞来完成这个任务。它们会对皮层中邻接区域的其他神经元产生强烈抑制作用，只允许一个“胜利者”出现。这些抑制性细胞只会影响到一个垂直柱的周边区域。因此，即使有大量抑制，一个区域中的多个垂直柱仍然能够同时兴奋。（在真正的大脑中，根本不可能只用一个神经元或一个垂直柱来表征事物。）为了方便理解，你可以假设一个区域中有且只有一个垂直柱激活。但你一定要记住，很多垂直柱可以同时产生兴奋。一个皮层区域对输入信息进行分类的过程，以及学习这种能力的过程，都是非常复杂的，至今尚未探索清楚。这里我并不想让你蹚这个浑水，我想假设我们的皮层区域已经将输入信息分类，

并转换为一组垂直柱的兴奋状态。接下来，我们就能关注序列的产生以及序列的“名称”上。

我们的皮层区域是如何存储这些分类后的模式序列的呢？前面我已经暗示过这个问题的答案，现在让我来进行更为详细的介绍。假设你是一个垂直柱，来自较低层区域的输入信息激活了你的第4层细胞。导致你第4层细胞兴奋。你很高兴，接着你的第4层细胞激活了第2、第3层细胞，然后是第5层，进而导致第6层细胞被激活。这样整个垂直柱就被低层区域输入信息激活了。第2、第3、第5层每一层中的细胞都有成千上万的突触在第1层。当其中一些突触随着第2、第3、第5层的激活而激活时，这些突触就会得到加强。如果这种情况发生足够多次，第1层的这些突触就会变得足够强，能够让第2、3、5层的细胞在第4层细胞没有激活的情况下也被激活——也就是说即使没有得到更低层区域的输入信息，垂直柱中的某些部分也能被激活。这样，第2、第3、第5层细胞就能根据第1层的模式预测应该何时激活。在这种学习之前，垂直柱只能被第4层细胞激活。而在这之后，垂直柱能够根据记忆部分地激活。当垂直柱通过第1层中的突触激活时，它就是在预期来自下方较低层区域的输入信息。这就是预测。如果垂直柱会说话，它会说：“过去当我被激活的时候，我第1层中的这部分突触也是激活的。所以，当我看到这部分突触再次被激活的时候，我就会预先激活。”

第1层接收的输入信息有一半来自相邻垂直柱和区域的第5层细胞。这些信息代表了刚刚发生的事件。它代表在你的垂直柱激活之前被激活的垂直柱。它就是旋律中前面的音程，刚刚看到的东西，刚刚感受到的东西，或者刚刚所听讲话的最后一个音素。如果这些模式发生的顺序总是不变的，那么垂直柱就能学到这个顺序。它们会按照顺序依次激活。

第1层所接受输入信息的另一半来自更高区域的第6层细胞。这些信息更加稳定。它代表你现在正经历的序列的名字。如果你的垂直柱是音程，那么它就是旋律名称；如果你的垂直柱是音素，那么它就是你听到

的那个词；如果你的垂直柱是口语单词，那么它就是你正在背诵的演讲。因此，第1层中的信息既表征序列的名字，也表征序列中的前一个模式。这样，多个不同序列就能分享同一个垂直柱，而不致引起混淆。垂直柱能够在正确的上下文中、以正确的顺序激活。

在介绍新的内容之前，我需要指出，第1层中的突触并不是唯一参与预测垂直柱何时激活的突触。我前面也讲了，垂直柱细胞会与周围的垂直柱互相传递信息，有超过90%的突触来自该垂直柱之外的细胞，而且这些突触中的大部分都不在第1层。例如，第2、第3、第5层的细胞会在第1层有成千上万个突触，但在它们自身层中也有着成千上万个突触。一个总的思路就是，细胞需要收集任何有用的信息来帮助自身预测何时会被来自下层区域的信息激活。通常，临近的垂直柱之间有较强的关联性，所以我们能看到很多连到邻近垂直柱的直接连接。例如，如果有一条线在你视野中移动，这会激活若干连续的垂直柱。但是一般而言，用来预测垂直柱激活的是更加全面的信息，这也是第1层突触的作用。如果你是一个细胞或一个垂直柱，你是不知道这些突触的作用的，你所知道的仅仅是它们能帮助你预测什么时候激活。

* * *

现在我们考虑下面这个问题，大脑皮层区域是如何为一个已习得的序列形成名字的呢？让我们再一次假设你是大脑皮层中的一个区域。你的活跃垂直柱随着每一个新的输入信息而不断变化。你已经成功学习到了垂直柱激活的次序，这表示在较低层区域的输入信息到来之前你的垂直柱中的一些细胞就开始兴奋了。那么，你会向更高层区域发送什么信息呢？我们前面看到，你的第2、第3层细胞会将轴突伸向更高区域。这些细胞的激活状态就是向较高区域发送的输入。但有个问题。为了使皮层体系有效工作，你必须在习得序列的发生期间向上发送恒定模式，也就是说，你要发送的是序列的名字，而不是细节。在你学习到一个序列之前，你可以发送细节，但当你习得这一序列并能成功与其垂直柱激活

后，你应该只发送恒定模式。然而，我还没向你展示过应该如何实现这一点。到目前为止，不管你是否能作出预测，你都会发送每个变化的模式。当每个垂直柱被激活时，它的第2、第3层都会发送一个新的信号到较高层区域。大脑皮层需要某种方式让发送到较高层区域的输入信息在已习得序列发生期间保持不变。当某个垂直柱能够作出预测时，我们需要某种方式将它的第2、第3层的输出模式关闭；而当该垂直柱无法预测的时候，仍能让第2、第3层细胞保持激活。这是实现恒定名称模式的唯一办法。

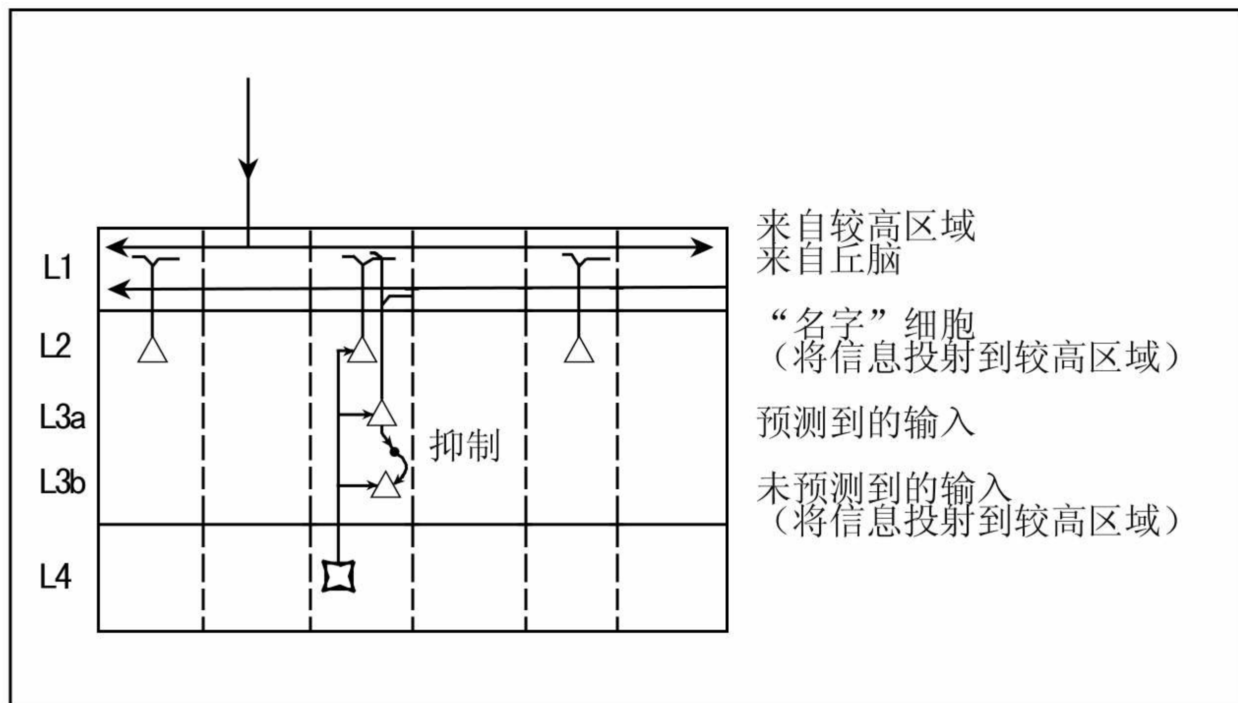


图10　为习得的序列形成一个名字

我们并不确切知道大脑皮层是如何做到这一点的。我能想象到几种方法。这里我将只介绍我最喜欢的一种，但是请记住，这个概念要比特定方法更重要。创造恒定名称模式是这个理论的必要条件之一。现在我将展示实现这一命名过程的最有可能的机制。

如图10所示，再次想象你自己是一个垂直柱。我们希望知道你是如何实现以下功能的：当你能预测到你的激活后，会把一个恒定模式发给更高层区域，而当你不能预测的时候，则会发送一个变化的模式。现

在，我们假设第2、第3层有几种不同类型的细胞。（除了几种抑制细胞之外，很多解剖学家还区分出两种层内的细胞，他们称之为3a层和3b层细胞，因此这个假设是有道理的。）

我们还假设其中一类细胞（称为第2层细胞）会在习得的序列中保持激活。这些细胞作为一个整体表征序列的名字。只要我们的皮层区域能够预测哪些垂直柱接下来会被激活，它们就会向更高区域传递一个恒定的模式。如果我们的区域习得的是一个含有3个模式的序列，那么只要我们还在这个序列中，所有表征这3个模式的垂直柱中的第2层细胞都将保持激活状态。它们就是这个序列的名字。

接下来，让我们假设有另外一类细胞（第3b层细胞），这些细胞会在垂直柱成功预测输入信息的时候不产生激活，而在垂直柱无法预测的时候产生激活。一个第3层细胞表征预期以外的模式。它会在垂直柱发生无法预测的兴奋时激活。当一个垂直柱没有学习之前，第3b层细胞经常激活。而当垂直柱学会对它的激活状态作出预测后，第3b层细胞就平静下来了。第2层细胞和第3b层细胞一起实现了我们的需求。在学习之前，这两种细胞都随着垂直柱的状态改变激活状态，而在学习之后，第2层细胞就会持续保持激活，而第3b层细胞则处于静息状态。

这些细胞是如何做到这一点的呢？首先让我们来看看当垂直柱成功地预测到自己的激活状态时，第3b层细胞是如何被抑制的。假定第3b层细胞之上还有另外一种细胞，叫作第3a层细胞。该细胞也有树突伸展到第1层。它唯一的作用是，当它在第1层看到适当的模式时，就会避免第3b层细胞被激活。当第3a层细胞看到第1层细胞中的已习得模式，它会快速激活一个抑制性细胞，从而避免第3b层细胞激活。这就是当垂直柱正确预测自己的激活状态时，它为了抑制第3b层细胞激活所做的一切。

现在来考虑一个更难的任务，即在整个已知模式序列中始终保持第2层细胞的兴奋状态。这个任务更难的原因主要是，很多不同垂直柱的多种第2层细胞需要共同保持兴奋，即使某些垂直柱本身并没有被激

活。我相信这是可能发生的，其原理可以这样来解释。第2层细胞能够学习完全只被来自皮层更高层区域的输入信息驱动。它们会优先与来自上层区域的第6层细胞的轴突形成突触。这样一来，第2层细胞就能够表征来自更高层区域的常量名字模式。当更高层区域向下层区域的第1层发送一个模式的时候，下层区域的一组第2层细胞会被激活，它们代表属于该序列的所有垂直柱。由于这些第2层细胞还会反过来向更高区域投射，因此它们会形成一个半稳定状态的细胞组。（这些细胞不太可能一直保持兴奋。它们更可能是按某种节奏同步激活。）就像更高层区域把音乐名称发送给下层区域的第1层，这会使一系列第2层细胞产生激活，其中每一个都对应着听到这个音乐时会被激活的垂直柱。

这些机制共同发挥作用，就能够让大脑皮层学习序列、作出预测、形成序列的常量表征（或者“名字”）。这都是形成恒定表征的基本操作。

对于过去未曾见过的事件，我们如何作出预测呢？在对一个输入信息的多种理解之间，我们如何作出选择呢？皮层区域如何基于恒定记忆作出某个预测？关于这些问题，前面我举过一些例子。比如，在你只能回忆出两个音符之间的音程时，去预测音乐中的下一个音符；还有关于火车的比喻，以及背诵葛底斯堡演讲等等。在这些情形下，解决问题的唯一办法是利用已出现信息将恒定预测转换成特定预测。从大脑皮层的角度来说就是，我们必须将前馈信息（真实输入）和反馈信息（恒定形式的预测）相结合。

这里给个简单的例子来说明这个过程。假设你的皮层区域被告知会听到一个五度音程。皮层区域的垂直柱能够表征出所有可能的特定音程，例如C-E调、C-G调和D-A调等。你需要决定应该激活哪些垂直柱。当更高层区域告诉你会有一个五度音程的时候，它会让所有表征五度音程（如C-G调、D-A调和E-B调）的垂直柱的第2层细胞激活，而表征其他音程的垂直柱的第2层细胞将不被激活。现在，你需要从所有表征五

度音程的垂直柱中选择一个。进入你区域的输入是一些特定的音符。如果你听到的上个音符是D调，那么所有包含D调的音程（如D-E调和D-B调）的垂直柱都得到了一部分的输入信息。因此现在在第2层上，所有表征五度音程的垂直柱都被激活了，而在第4层上，所有表征包含D调的音程的垂直柱也得到了部分输入信息。这两者之间的交集就代表了我们的答案，即表示D-A调音程的垂直柱（如图11所示）。

大脑皮层是如何找到这个交集的呢？前面我曾提到过，第2、3层细胞的轴突在离开大脑皮层的时候通常会在第5层形成突触，而从较低区域到第4层的轴突也会在第6层形成突触。这两种突触（自上而下和自下而上）的交集正好满足了我们的需要。同时接收到这两种输入信息的一个第6层细胞会被激活。第6层细胞表征着一个皮层区域认为正在发生的事情，也就是某个特定的预测。如果第6层细胞会说话，它会说："我是表征某个事物的垂直柱的一部分。以我为例，我所在的垂直柱表征着音程D-A调。其他垂直柱表征其他的事物。我为我所在的大脑皮层区域代言。当我被激活的时候，就说明我们认为已经出现或者即将出现音程D-A调。我之所以被激活，有可能是由于来自耳朵的自下而上的输入信号导致我的垂直柱第4层细胞激活，从而激活了整个垂直柱。也有可能意味着我识别出一段旋律，预测到D-A调就是下一个音程。无论是哪一种可能，我的任务都是告诉较低层的皮层区域我们认为正在发生的事情。我代表了我们对这个世界的理解，无论那是真实的还是想象的。"

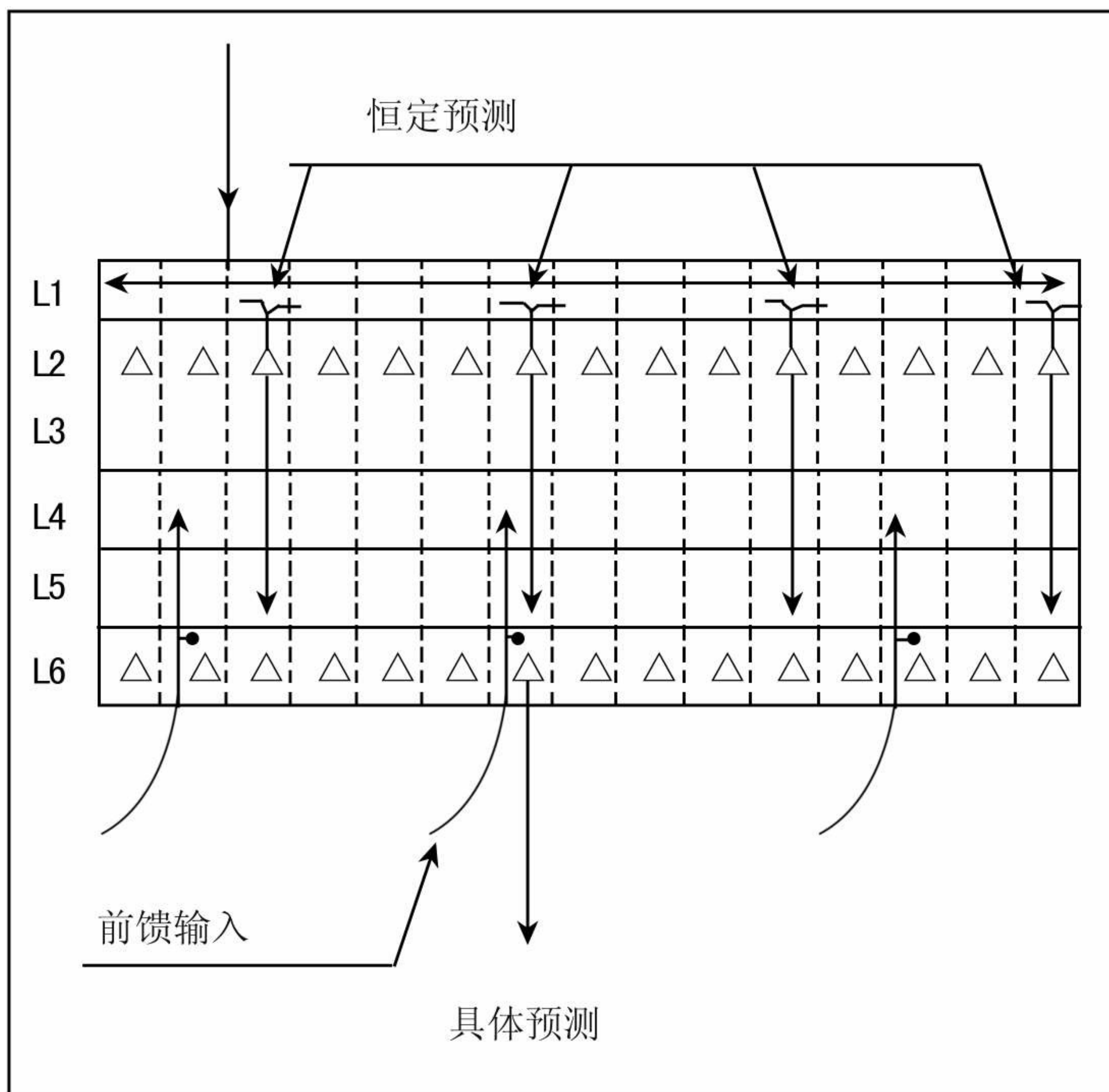


图11　一个皮层区域根据恒定记忆作出某个预测

下面我们用另外一种心理图景来描绘这一机制。假设有两张纸，上面打有很多小洞。其中一张纸上的洞代表那些第2层或第3层细胞被激活的垂直柱，也就是我们的恒定预测。另外一张纸上的洞代表那些有低层区域输入信息的垂直柱。如果我们将两张纸重叠在一起，有一些洞会重合，而另一些则不会。重合的洞就代表着我们认为需要激活的垂直柱。

这一机制不仅能作出具体的预测，还能对感官输入的信息消歧。一

个皮层区域的输入信息经常是带有歧义的，例如我们举过的彩色纸的例子，或者听到的模棱两可的词。这种自底向上、自上而下的匹配机制能够让你在两种或多种可能的理解中作出选择。一旦决定了，你就会将你的理解传给下层区域。

在你人生清醒的每时每刻，大脑皮层的每个区域都在不断比较上层驱动的一组预测垂直柱和下层驱动的一组观测垂直柱。两组的交集就是我们所感知到的事物。如果我们有来自下层的完美输入和来自上层的完美预测，那么我们的感知垂直柱集合将总是预测垂直柱集合的子集。但并不是总能达成一致。这种将部分预测和部分输入信息结合的方式，可以消除输入信息的歧义、补充丢失信息，在不同的理解之间作出选择。我们就是通过这种方式将预测的恒定音高的音程与听到的最后一个音符相结合，来预测旋律中的下一个音符的。我们也是通过这种方式来判断一张图片到底是一个花瓶还是两张人脸的。同样也是通过这种方式，我们能够将运动指令流分开，分别写出和朗诵葛底斯堡演讲。

最后，除了向较低层皮层区域投射信息，第6层细胞还能够将输出信息送回到所在垂直柱的第4层细胞。如此一来，我们的预测就成了输入信息。这就是当我们在做白日梦和思考的时候所用到的皮层机制，它能让我们看到自己预测的结果。我们每天都会花几个小时做此类事情——计划未来、排练演讲以及担心即将到来的事。长期从事大脑皮层模型研究的斯蒂芬·罗斯伯格称该现象为“折叠反馈（folded feedback）”。我更喜欢称之为“想象”。

* * *

这是该章节的最后一个话题。我曾反复指出过，我们所见、所闻和所感经常高度依赖于我们的自身行动。我们能看到什么，取决于我们眼睛扫视哪里以及我们的头如何转向。我们能感受到什么，取决于我们如何移动四肢和手指。我们能听到什么，有时则取决于我们说了什么、做了什么。

因此，要想预测接下来我们将感知到什么，我们必须知道我们正在做什么动作。运动行为和感官知觉相互之间是高度依赖的。如果我们接下来的感知很大程度上取决于自身行动的话，我们将如何作出预测呢？幸运的是，我们对这个问题有个惊人且简洁的解决方案，虽然其中有很多细节尚不清楚。

第一个惊人发现是，我们的知觉和行动几乎是统一的。如前所述，几乎所有皮层，甚至包括视觉区，都会参与运动的产生过程。投射到丘脑和第1层的第5层细胞，似乎也参与运动功能，因为它们会同时将信息投射到旧脑的运动区。因此，关于“刚刚发生了什么”的信息，无论是知觉的还是运动的，都能传递到第1层细胞。

第二个惊人发现是，运动行为也必须表征为层级的恒定表征，这是第一个事实的必然结果。为了完成某个特定动作，你需要产生一系列运动，而这需要你的大脑通过“细节-恒定”的形式来产生。在行动指令沿着层级向下传递的过程中，它会被转化成更为复杂而细致的序列，从而实现你期望发生的运动。这在“运动”皮层和“知觉”皮层都有发生，从而混淆了两者的区别。假设视觉皮层的IT区感知到了“鼻子”，如果想要切换到“眼睛”的表征，就会产生必要的扫视来将预测变为实现。从鼻子移动到眼睛的扫视取决于脸的位置。近处的脸需要更大幅度的扫视，远处的脸只需要小幅度的扫视。一张倾斜的脸和一张水平的脸所需的扫视角度亦不相同。扫视的细节会随着对“眼睛”的预测不断流向V1区而逐级确定。随着预测信息不断向下流动，扫视变得越来越确定，从而让视网膜中央凹恰好对准或非常接近目标。

再举个例子。假设我要从卧室走到厨房，我的大脑需要做的是将卧室的恒定表征转换到厨房的恒定表征。这一转换会带来复杂的序列展开过程。在这个过程中，会不断产生对从卧室走到厨房过程中可能看到、感到和听到之事物的预测序列，同时也会产生运动指令序列，能让我走过去，并在这个过程中不断移动眼睛。在模式上下流动的过程中，预测

和运动行为协同工作着。当你行动的时候，你的预测不仅先于感官，还决定着感官，这听上去很神奇。在序列中，关于移动到下个模式的想法，会产生对接下来发生事情的级联式（cascading）预测。当预测展开后，就会产生必要的运动指令来实现预测。思考、预测以及行动都是随大脑皮层逐级向下流动的同一个序列展开过程的一部分。

思而行，对于知觉和运动行为的同步展开，是所谓“目标导向”行为的本质。目标导向行为是机器人学的“圣杯”。它内置于大脑皮层中。

我们当然可以停止我们的运动行为。我可以想象看到了什么，而不必真的看到它。我也可以想象去了厨房，而不必真的去。但是，“想做某事”的确是我们真正去做的开始。

向上流动和向下流动（Flowing up and Flowing Down）

让我们稍微回顾一下，信息是如何在大脑皮层体系中上下流动的。当我们在这个世界里运动时，不断变化的输入信息流进入大脑皮层的较低层区域。每个区域都试图将它得到的输入信息流解读为已知模式序列的一部分。区域的各垂直柱试图预测它们的激活行为。如果预测成功，它们会将一个稳定模式，即序列名称，传递到更高区域。这就好像区域在说：“我在听一首歌，这就是它的名字。我可以处理这些细节。”

但如果来了一个预料之外的模式，例如一个想不到的音符，那该怎么办呢？或者，如果我们看到了一个并不属于脸的东西，会怎么样？预测之外的模式会被自动传递到更高层区域。当不属于可预测序列的那些第3b层细胞激活的时候，这很自然就会发生。更高层区域也许能够将这个新模式作为它的已知序列的下一个部分来理解。就好像是说：“啊，我听到一个新的音符。这也许是这个专辑中下首歌的起始音符。看来是

这样，所以我预测我们已经跳到下首歌了。喂，下面的区域，这就是下首歌的名字，你应该就在听这首歌。”但如果这个模式一直没有被识别，它就会不断被向上传递，直到更高区域能够将其理解为自己正常序列的一部分为止。这种模式所到区域层次越高，参与解决这个预测之外的输入信息的区域就越多。最终，当某个上层区域认为它能解读这个事件了，它就会产生一个新的预测。这个新的预测沿着皮层逐级向下传播。如果这个新的预测不准确，错误就会被检出，这样它又会再次沿着大脑皮层逐级向上传递，直到有区域可以正确理解它。因此我们可以看到，观测到的模式会沿着大脑皮层逐级向上流动，而预测会逐级向下流动。理想情况下，在一个已知和可预测的世界中，大部分上下流动的模式会快速发生，并且主要出现在大脑皮层的较低层区域。大脑会迅速地尝试找到与预测之外的输入信息中相匹配的那部分世界模型。只有这样它才能理解输入信息，并知道接下来会发生什么。

如果我在家中的一间熟悉的房间里走动，在大脑皮层中低层区域几乎不会有错误向上层扩散。关于我家的序列已经被熟练掌握了，在视觉、躯体感觉和运动系统的低层区域就能得到很好的处理。我太了解这个房间，甚至可以在黑暗中来回走动。对周围世界的熟识能够解放大部分大脑皮层，可以让我进行其他任务，例如思考大脑结构、写书。但假设我是在一个陌生的房间里，尤其是如果它有别于我见过的其他任何房间，那么我不但要观察自己走到哪儿了，而且预测外的模式会不断沿着大脑皮层向上传递。我的感官体验越是无法与已经习得的序列匹配，就会有更多的错误被向上传递。在这种情况下，我就再也不能专心思考大脑了，因为我的大部分大脑皮层都忙着参与房间探路了。这是那些走下飞机踏上异国土地的人们所共有的体验。尽管公路看上去可能与之前所熟悉的没什么两样，但汽车可能是在路的另一边行驶；货币是陌生的，语言也是很难理解的，甚至连找个卫生间都需要动用你全部的大脑资源。走在异国的路上，你是无暇排练演讲的。

所谓顿悟的感觉，也就是“啊哈，灵机一动”的时刻，也能用这个模

型解释。假设你正在看一张模糊不清的图片。图片上布满了杂乱的墨点和线，看不出像任何东西，似乎毫无意义可言。当大脑皮层发现任何记忆都没法匹配输入信息的时候，就会产生困惑。你的眼睛会扫描图片的每个位置。新的输入信息会通过所有通路沿着大脑皮层向上流动。高层皮层会尝试各种不同假设，但是当这些预测向下流动时，它们又会与输入信息相冲突，大脑皮层不得不重新尝试其他的可能性。在此期间，你的大脑完全投入到理解这张图片上来。最终，你作出了一个准确的高层预测。然后，这个预测会从大脑皮层的顶层开始向下传递。在不到一秒钟的时间里，每个区域都会得到与输入数据匹配的序列。不再有错误向上传递。如图12所示，你理解了这张图片，你在这些杂乱的墨点和线段之间看到了一只斑点狗。



图12　你看到那只斑点狗了吗

反馈真的能起作用吗？

几十年来，我们一直都知道大脑皮层中的连接是相互的。如果区域A投射到区域B，那么区域B也会投射到区域A。通常情况下，反馈的轴

突纤维要比前馈的多。然而，尽管这样的描述广为人们所接受，人们却普遍认为，反馈在大脑中只起到辅助或调解作用。认为反馈信号能够迅速准确地引起第2层中多组细胞激活的看法，并不为神经科学家普遍接受。

为什么会这样呢？如前所述，部分原因在于，如果你不接受“预测”的核心作用，那么也就没有必要去考虑“反馈”了。如果你认为信息是直接流向运动系统的，那为什么还需要反馈呢？忽视反馈的另一个原因在于，反馈信号遍及第1层的广大区域。直观上，我们认为一个散布于较大区域的信号对许多神经元只会起到很小的作用，而大脑中确实存在很多种这样的调节信号，它们不对特定神经元起作用，而只是改变诸如警觉性一类的全局属性。

忽视反馈的最后一个原因在于，没有多少科学家相信神经元个体会发挥作用。典型的神经元往往有成千上万个突触。这些突触有的远离细胞体，有的则非常靠近细胞体。距离细胞体较近的突触对细胞的激活有较强影响。十几个距离细胞体较近的兴奋突触就能让神经元产生动作电位。这是众所周知的。但是，绝大多数突触离细胞体很远。它们广泛分布在像树枝一样的树突上。由于这些突触远离细胞体，科学家倾向于认为，这些突触上的动作电位对其所在神经元的影响微乎其微。远距离突触的作用会随着它到达细胞体的过程而消减。

作为普遍规律，沿着大脑皮层向上流动的信息是通过细胞体附近的突触传递的。因此，向上流动的信息在传递过程中越来越确定。同样作为普遍规律，沿着皮层向下流动的反馈是通过细胞体远处的突触传递的。在第2、3、5层的细胞会将树突伸向第1层，并在那儿形成突触。第1层中包含大量突触，但都远离第2、3、5层的细胞体。而且，第2层中的任意一个细胞与任一特定的反馈纤维之间，即便会形成突触，数目也极少。因此，对于第1层中的简要模式会准确引起第2、3、5层细胞激活的观点，有些科学家持反对意见。而这个观点正是我的理论所必需的。

解决这一困境的一个办法是，假设神经元采用的是与传统模型不同的工作模式。实际上，近年来已经有越来越多科学家提出，远距离的细树突上的突触能够在细胞激活中扮演积极且具有高度特异性的角色。在这些模型中，远距离突触的作用与近距离的粗树突上的突触不同。例如，如果在一个细树突上有两个特别接近的突触，它们就会形成“重合检测器（coincidence detector）”。也就是说，如果两个突触在短时间内都收到动作电位输入信息，那么即便它们距离细胞体很远，也会对该细胞产生很大的影响。它们会让细胞体也产生一个动作电位。神经元的树突究竟是如何工作的，这仍然是个谜，所以这里我不会说太多关于树突的事情。但重要的是，大脑皮层的记忆-预测模型需要远距离的突触能够检测特定的模式。

回头来看，认为神经元上数以千计的突触只发挥调节作用的看法很不明智。大量反馈信息和突触的存在一定有它的道理。现在，我们可以说，一个典型的神经元通过在细树突上形成突触，使得自己能够学习、反馈纤维上的成千上万个精确的重合。这意味着，神经皮层中的每个垂直柱都高度灵活，能够被不同的反馈模式激活。这也意味着，任何特定的特征都能够与成千上万个不同对象和序列产生精确的关联。我的模型要求反馈又快又准。当细胞看到远距离树突上的任何数目的精确重合，它都要激活。在新的神经元模型中，这一点是可以被满足的。

大脑皮层如何学习？

大脑皮层所有层次上的细胞都有突触，其中大多数突触都可以通过经验修改。可以说，学习和记忆发生在所有的层次、所有的垂直柱和所有的分区中。

前面我们提到过以加拿大认知心理学家唐纳德·赫布（Donald O. Hebb）的名字命名的“赫布学习法”。赫布学习法的要点很简单：当两个

神经元同时激活，它们之间的突触就会得到增强（这可以简单总结为“共同激活，共同连接”）。我们现在知道赫布基本上是正确的。当然，自然界没有什么是简单的，在真正的大脑中会有更多复杂的细节。我们的神经系统中有很多赫布学习法规则的变种。例如，有些突触会根据神经信号时长的细微变化而改变强度，有的突触变化是短时的，有的则是长时的。赫布只是为学习行为建立了一个研究框架，并非最终理论，而这个框架已经非常有用了。

赫布学习法理论可以解释本章涉及的大部分大脑皮层行为。请记住，早在19世纪70年代，采用经典赫布学习算法的自-联想记忆（auto-associative memories）就能够学习空间模式和模式序列了。可是问题在于它无法很好地处理变化。根据本书提出的理论，大脑皮层可以通过两种机制解决这个问题，一是将自-联想记忆建立层次结构，二是采用更为复杂的垂直柱结构。本章主要阐述的便是这一层次结构及其工作原理，因为正是这一结构让大脑皮层格外强大。因此我着重于介绍有关层次结构学习的一些主要原则，而不是去探讨每个细胞如何学这学那等一些令人头疼的细节。

当你刚刚出生，你的大脑皮层基本上什么也不知道，对母语、文化、家庭、家乡、歌曲以及将要伴随你成长的亲人，它都一概不知。所有的这些信息，以及这个世界的结构，都需要你学习。学习包括两个基本模块：形成模式分类和构建模式序列。这两个记忆模块互为补充，并且相互作用。当某个区域学习序列时，它传送到更高区域第4层细胞的输入就会不断变化。这些第4层细胞因此会学习形成新的分类，从而改变投射回更低区域的第1层细胞的模式，这反过来又会影响序列学习。

形成序列的基本思想是将有关同一对象的模式聚为一组。其中一种方法是将时间上连续出现的模式聚为一组。如果一个小孩手里拿着玩具并且慢慢地移动它，她的大脑就会认为视网膜上的图像是关于同一个对象的，因此这些变化的模式就会被聚在一起。而在另外一些情况下，你

需要借助外部指令来帮助你判断哪些模式应当聚在一起。这就像是，你需要借助老师的指导，才能了解苹果和香蕉都是水果，而胡萝卜和芹菜不是。无论使用哪种方式，你的大脑都会逐渐建立起属于一个整体的模式序列。但随着大脑皮层的某个区域建立起序列，它送到下个区域的输入信息就会发生变化。输入从主要表征独立的模式变成表征模式组合。区域的输入从音符变成旋律，从字母变成单词，从鼻子变成人脸，如此种种。随着自下而上的区域输入信息越来越“面向对象”，高层区域就可以学习更高阶的对象的序列。前面的区域建立起字母序列，这个区域就可以建立单词序列。这个学习过程产生了出人意料的结果：通过反复学习，对象的表征沿着大脑皮层逐级下移。在你人生的最初几年里，对于世界的记忆最先形成于大脑皮层的较高区域，随着学习，它们会在越来越低的皮层区域中重构。大脑并没有移动它们，而是在一遍又一遍地重新学习它们。（我并不是说所有记忆都起始于大脑皮层的最高层。记忆实际的形成过程要复杂得多。我认为第4层的模式分类起始于较低层区域，然后逐渐上移到较高层。但与此同时，序列开始在较高层形成，并逐渐下移。也就是说，是序列记忆不断在较低大脑皮层区域中重构。）随着简单表征的下移，高层区域就能够学习更复杂而细致的模式了。

你可以通过观察儿童的学习来发现层次记忆的构建和下移。想想我们是如何学习阅读的。我们首先学习的是识别单个的印刷体字母。这个任务漫长而艰难，需要有意识地付出努力。然后，我们开始认识简单的单词，一开始这也很漫长而艰难，即使是学习一些3个字母的英文单词。儿童能按顺序认出单词中的每个字母并读出来，但要经过大量练习之后才能认识整个单词。简单的单词之后，我们开始挑战更复杂的多音节单词。最初，我们会读出每个音节，像上面对待字母那样把它们连起来。经过多年练习，我们就可以快速阅读了。这时，我们不再留意单个字母，而是一瞥之间就能认出整个单词甚至短语。这不仅仅是因为我们变快了，还因为我们实际上会将单词和短语当作一个整体来识别。当读到一个单词时，我们是不是还会看每个字母呢？答案既是也否。显然，视网膜会看到所有字母，视觉皮层的V1区也能看到。但是对字母的识

别发生在大脑皮层的较低层区域，例如V2和V4区。当信号进入到IT区时，就不再有对字母的表征了。最初需要你动用整个视皮层来完成的任务——识别单个字母，现在在感官输入信息之后就立即完成了。随着对字母等简单对象的记忆的下移，更高层区域就有能力学习如单词和短语等更为复杂的对象了。

学习识乐谱也是这样的。最初你必须全神贯注于每个音符。随着不断训练，你开始熟知常见的音符序列，乃至整个乐句。在大量训练之后，你就好像看不到大部分音符了。摆在那儿的乐谱仅仅是帮你想起这部乐曲的结构，更细节的序列记忆已经存储在了较低区域中。在运动和知觉领域都有这样的学习。

儿童的大脑无论在识别输入方面还是在做出运动指令方面都比较慢，这是因为完成任务所需的记忆位于大脑皮层的较高区域。为了解决冲突，信息不得不上下流动很多层级，甚至需要往返多次。神经信号沿着大脑皮层的上下流动是需要花费时间的。儿童的大脑也尚未在高层形成复杂序列，因此无法识别或回放复杂模式。儿童的大脑还无法理解这个世界的高阶结构。与成人相比，儿童的语言、音乐和社会交互都很简单。

如果你反复钻研某一组特定的对象，你的大脑皮层会重构对这些对象的记忆表征，并逐级下移。这就可以把较高层的大脑皮层空闲出来，去学习更为细致、复杂的关系。根据该理论，专家就是这样产生的。

在我的计算机设计工作中，许多人对我能够一眼就看出产品的设计问题感到非常惊讶。经过了25年的计算机设计历练，在移动计算设备的相关问题上，我已经建立了一个高于平均水平的模型。类似地，有经验的父母会很容易知道他们的孩子为什么情绪低落，而年轻父母则要为这个问题颇费脑筋。有经验的业务经理可以很容易发现某个组织结构的优缺点，而新手经理则很难理解这些。虽然他们的输入信息都相同，但是新手们的模型不够精致复杂。在所有类似的情形中，我们都是从最基

础、最简单的结构开始学起的。随着我们的知识沿大脑皮层逐级下移，我们就有机会在较高区域学习高阶结构了。正是这些高阶结构让我们富有经验。专家和天才的大脑能够看到结构背后的结构，模式背后的模式，这是普通人看不到的。你可以通过不断训练成为专家，但对天赋和天才而言，当然也含有遗传的成分。

海马体：在最顶层（The Hippocampus: on Top of it all）

在大脑皮层下方，有3个能与之交流的脑结构。它们分别是基底核（basal ganglia）、小脑（cerebellum）和海马体（hippocampus）。这3个结构先于大脑皮层而存在。粗略来讲，基底核是原始的运动系统，小脑负责学习事件间精确的时间关系，而海马体则存储与特定事件与地点有关的记忆。从某种程度上说，大脑皮层囊括了这些结构的原有功能。例如，一个生来就缺损小脑的人，将会在计时方面存在缺陷，因此需要在移动时非常留神，但在其他方面则很正常。

我们知道大脑皮层负责所有的复杂运动，能够直接控制你的四肢。但这并不是说基底核不重要，只是说大脑皮层接管了大部分的运动控制功能。因此，我在前面介绍大脑皮层的全部功能时，是将其独立于基底核和小脑的。有些科学家可能不同意这种假设，但这本书和我的研究工作都采纳了这一假设。

海马体则很不同。它是大脑中被研究最多的区域，因为它对新记忆的形成至关重要。如果你同时失去了左右两半海马体（就像神经系统的很多部分那样，海马体同时存在于大脑的左右两侧），你就无法再形成新的记忆。如果没有了海马体，你仍能说话、走路和听到声音，短期内看起来就跟正常人一样。但实际上你已经受到了严重的损伤：你无法记住任何新的东西。你能记得失去海马体前认识的朋友，但却无法记住新

认识的人。即使一年中你会与医生见5次面，对你而言，每次见面都会像是初次见面一样。你对失去海马体后所发生的事情毫无记忆。

多年来，我一直不愿去思考海马体的作用，因为它对于我的理论而言毫无道理可言、无法理解。很明显它对学习很重要，但它又不是储存我们所知道的大多数事情的终极知识库，大脑皮层才是。对海马体的传统观点认为，新的记忆在海马体中形成，然后经过几天或几个月之后，这些新的记忆会被传输到大脑皮层中。我认为这毫无道理可言。我们都知道景象、声音、触觉——这些感官数据流，都是直接流入到大脑皮层的感觉区，并没有先流经海马体。在我看来，这些感觉信息应该在大脑皮层中自动形成新的记忆。我们为什么需要海马体来学习呢？像海马体这样一种独立的结构，怎么就能干预甚至阻碍大脑皮层中的学习行为呢？它又如何将这些信息传回到大脑皮层呢？

我决定将海马体放在一边，心想总有一天它的作用会被搞清楚。这一天发生在2002年年底，正是我开始写这本书的时间。我的一个在红杉神经科学研究所（Redwood Neuroscience Institute）的同事布鲁诺·奥尔斯豪森（Bruno Olshausen）指出，海马体和大脑皮层之间的连接表明，海马体是大脑皮层的顶层区域，而不是一个独立的结构。该观点认为，海马体处在大脑皮层金字塔的顶端，也就是图5的最顶部。在进化过程中，大脑皮层出现在海马体和其他大脑结构之间。很显然，这一观点的存在已经有一段时间了，只是我还不知道而已。我与几位海马体专家探讨过，并让他们解释为什么这个海马形状的结构能够将记忆传给大脑皮层。没人能解释，也没人告诉我海马体处在大脑皮层金字塔的顶端，这也许是因为海马体不仅位于大脑皮层金字塔的顶端，却也跟其他的大脑旧部直接相连。

但我很快就意识到，这一新的视角正是解开我当前困惑的关键所在。

想想从你的眼睛、耳朵和皮肤流入大脑皮层的信息。大脑皮层的每

个区域都尝试理解这些信息的意义。每个区域都试图将输入理解为它所知道的序列。如果理解了输入，它就会说：“我知道这个，它只不过是我已经看到的对象的一部分。我就不再向上传递这些细节了。”如果一个区域无法理解当前的输入信息，它就会将其向上传递，直到有更高层区域能理解它。然而，一个完全新奇的模式会向上逐步传递到最高层。每个更高层区域都会说：“我不知道这是什么，我解读不了，上面的兄弟来看看吧？”这带来的结果就是，到达大脑皮层金字塔顶的时候，就只剩下根据先前经验无法理解的信息了。这些全都是输入信息中完全新奇和出乎意料的信息。

每天我们都会遇到许多直达皮层顶端的新事物，例如报纸上的故事、早晨偶遇的人的名字，以及回家路上目击的车祸。正是这些无法解释和无法预期的事物会进入海马体并储存起来。但这些信息不会永远被保存在那里。它们要么被转移到大脑皮层中去，要么就彻底地丢失了。

我曾注意到，随着年龄的增长，我在记住新事物方面遇到了麻烦。例如，我的孩子能记得去年他们看过的大部分戏剧的细节，而我却不能。也许是因为我看过太多的戏剧，因此很难再有什么新鲜感。新戏的情节与老戏的记忆有雷同，因此这些信息就是到不了我的海马体。而对于我的孩子们而言，每部戏都是新的，都会传到海马体。如果这是真的，我们就可以说，你知道的越多，你记住的就会越少。

与大脑皮层不同，海马体的结构是异质的，有若干专门化的区域。它擅长于快速存储所看到的任何模式。海马体占据了能记住全新事物的最佳位置——大脑皮层金字塔的顶端。这也是回忆起这些全新记忆的完美位置，能够将它们转移并存储到大脑皮层的层级结构中去，当然这是一个相对较慢的过程。这样，你可以通过海马体很快地记住一个全新事件，但当你反复经历或思考某个事件时，你就会在大脑皮层中永远地记住它。

通往大脑皮层顶层的另一条通路

你的大脑皮层中还有另外一条主要通路，能让信息从一个到另一个区域逐级向上传递。该通路始于第5层细胞，它会投射到丘脑（这是丘脑的另外一个部分，与我们前面提到的那部分不同），然后从丘脑投射到紧邻的更高层区域。每当两个区域之间存在层级直接连接时，它们也通过丘脑间接相连。这条第二通路只向上传递信息，而不向下传递。因此，当信息沿着大脑皮层体系向上传递时，会有一个连接上下区域的直接通路和一个流经丘脑的间接通路。

第二通路有着由丘脑细胞控制的两种操作模式。一种模式下，通路几乎是关闭的，因此信息无法通过。另外一种模式下，信息能够在区域间精确地流动。两位科学家——纽约州立大学石溪分校（State University of New York at Stony Brook）的穆雷·谢尔曼（Murray Sherman）和来自威斯康辛星大学医学院（University of Wisconsin School of Medicine）的雷·吉耶里（Ray Guillery），描述了该第二通路，并认为它可能与被本章作为主题讨论的直接通路同样重要（甚至可能更重要）。我对间接通路的功能有自己的思考。

请读一下这个词：想象（imagination）。大多数人一瞥之下就能认出这个词。现在看这个单词中间的字母i。然后看字母i上的那个点。你的眼睛一直注视着相同位置，但第一次你看到的是整个单词，第二次你看到的是一个字母，而最后一次你看到的是一个点。请盯住字母i，然后尝试将你的感知在单词、字母和点之间切换。如果这样做有困难，请你试着在盯着点的同时喊“点”、“i”以及“imagination”。在每一种情况下，进入V1区的信息都是完全相同的，但是当信息到达更高区域如IT区，你会感受到不同的事物、不同层次的细节。IT区知道如何识别这3种对象。它能识别出单独的点、字母i以及整个单词。但是，当你感知整个单词时，V4、V2和V1区负责处理细节，IT区仅知道这个单词本身。在阅读的时候，你通常不会感知到每个字母，你只会感知到单词或

者短语。但如果你想要感知字母，你也能够做到。我们每时每刻都在进行这样的注意力转移，但我们通常不会觉察到。我可以在听背景音乐时完全不注意它的旋律，但如果我想要的话，我可以从中分离出歌手或电吉他的声音。进入我头脑中的是同样的声音，但我可以将注意力聚焦到其中某个方面。你每次挠头时，手指的运动都会在头的内部产生很大的声音，但通常你觉察不到这个噪音。而如果你将注意力集中在这上面，就能清楚地听到它。感官输入的信息通常是在较低层大脑皮层被加以处理，但如果你关注它，它就会被带入更高层区域处理，上面就是一个例子。

我推测，这个经由丘脑的间接通路，就是当我们在注意那些通常不会注意到的细节时所用到的机制。它绕开第2层对序列的分组，直接将原始数据传给邻近的高层区域。生物学家已经表明，这个间接通路可以用两种方式打开。一种是通过来自更高层区域发出的信号。当我让你关注你通常不会关注的细节时，例如字母i上的点或者挠头的声音，你就会采用这种方式打开间接通路。第二种方式是来自下层的强大且在预料之外的信号。如果间接通路的输入足够强大，通路就会给更高区域发送唤醒信号，这样也可以打开通路。例如，如果我给你看一张脸，然后问你这是什么，你会说“脸”。但如果我给你看同一张脸，但鼻子上有个奇怪的标志，你会首先认出这张脸，然后你的较低层的视觉区域立刻就注意到有什么不对的地方。这个错误信号会迫使间接通路打开。相关细节会取道间接通路，绕过通常要进行的序列分组，然后你的注意力就会被吸引到这个标记上。这样，你就会看到这个标志，而不仅仅是脸了。如果这个标志足够奇怪，它就会吸引你全部的注意力。不寻常的事件就是通过这种方式迅速引起你的注意的。这也是为什么我们无法避免地去关注畸形或其他异常模式。你的大脑会自动完成这一切。然而，错误经常不会大到足以打开间接通路。这也是为什么我们有时没法发现阅读时的拼写错误。

结论

为了发现并建立新的科学理论框架，我们有必要去寻找最最简单的概念。这一概念要能够整合并解释大量的表面看来毫无联系的事实。这个过程带来的不可避免的后果是：由于过度简化导致偏离事实太远。重要的细节可能会被忽略，有些事实可能会被误解。但只要框架能站得住脚，就必然能发现其需要优化和修正的地方，比如知道最初的假设走得太远，或走得不够远，甚至从根本上就是错的。

在本章，我介绍了很多关于大脑皮层工作方式的设想。我觉得，其中一些想法最终可能会被证明是错的，甚至所有的想法都需要改进。此外，我还有很多细节没有机会提及。大脑太复杂了，如果是神经科学家在读这本书，他们就会知道，我不过是对真正大脑的复杂性作了一个非常粗略的描述。但我认为，这个框架从整体上是合理的。我所期望的是，虽然框架的细节会随着新数据和新理解的出现而发生变化，但其核心思想能被保留下来。

最后，你也许还是不相信，一个简单但庞大的记忆系统怎么就能实现人们所做的这么多事情。你我真的就是一个多层级记忆系统吗？我们的生命、信念和追求都能被存储在数万亿个微小的突触中？1984年，我开始从事专业计算机编程。之前我也写过一些小程序，但这次是我第一次用图形用户界面编程，也是我第一次编写复杂的大型应用程序。我是为Grid系统公司的一款操作系统编写软件。Grid操作系统在当时是非常先进的，拥有图形窗口、多种字体和菜单。

有一天，我对自己正在做的几乎不可能完成的任务感到了震惊。作为程序员，我每次只能输入一行代码。我将若干行代码放在一起，称为子程序。子程序又组成模块。模块最终组成整个应用软件。我当时所编写的电子表格软件包含了太多子程序和模块，以至于没人能读懂它。它太复杂了。而其中每一行代码的作用非常微小。在显示器上显示一个像

素就需要好几行代码，而要在整个屏幕上画满电子表格则需要计算机运行分散于几百个子程序中的上百万行指令。子程序又会重复或递归调用其他子程序。程序太复杂了，当它开始运行时，我们几乎不可能知道将会发生的一切。而当它真的运行起来，瞬间就绘制出了所需要的图形，这简直让人难以置信。绘制的图形看上去是包含数字、标签、文本和图示的表格，就像普通的表格一样。但我知道计算机内部不过是处理器在不断运行一个个简单的指令。很难相信计算机能够穿过模块和子程序的迷宫，如此快地运行所有指令。如果不是我事先有所了解，我一定会认为这不可行。我意识到，如果有人发明了一个带有图形用户界面和电子表格软件的计算机，并通过一张纸向我展示，我一定会认为这不切实际并加以拒绝。我一定会说这永远做不出来。这个想法让我感到羞愧，因为这些被证明是可能的。就是在那时，我意识到，对于微处理器的速度以及层级设计的性能而言，我的直觉估计是不充分的。

这对我们关于大脑皮层的看法来说是个教训。大脑皮层并没有运行速度超级快的模块，它的运行规则也并不复杂。但是，它拥有层级结构，其中包含数十亿个神经元，数万亿个突触。对于这样一个逻辑简单、数目庞大的记忆系统，如果说我们很难想象它是如何产生我们的意识、语言、文化、艺术、这本书以及我们的科学技术的话，我觉得这是因为我们的直觉对大脑皮层的容量和层级结构的性能估计不足。大脑皮层就是做到了。这不是什么魔法。我们一定能够理解它。正如我们发明了计算机，最终我们也能构建出拥有与大脑皮层相同工作原理的智能机器。

第七章 意识与创造力

在我报告我的大脑理论的时候，由于“预测”与大量的人类活动有关，听众总能够很快领会“预测”的关键性所在。他们会问很多与之有关的问题：创造力从哪里来？意识是什么？想象力是什么？我们如何把现实与错误信念区分开来？尽管这些话题并非我研究大脑的初衷，但它们确实是几乎所有人都会感兴趣的问题。我并不想假扮成这些问题的专家，但智能的记忆-预测框架能够为此提供一些答案和有益见解。在这一章，我将讨论几个最常见的问题。

动物有智能吗？

老鼠有智能吗？猫有智能吗？在演化历程中智能是何时出现的？我喜欢这样的问题，因为我发现它的答案令人惊讶。

本书写到目前为止，关于大脑皮层及其工作方式的一切论述都依赖于一个基本前提——世界是有结构的，因此是可预测的。世界包含许多模式：人脸上长眼睛，眼睛里有瞳孔，火是热的，重力会让物体下落，门可以开关，如此种种。世界既不是随机的，也不是同质的。如果世界没有结构，记忆、预测和行为都会变得毫无意义。所有的行为，无论是人的、蜗牛的、单细胞生物的、还是一棵树的，都是在利用这个世界的结构进行繁衍。

想象一下池塘里的一个单细胞动物。它长有一条鞭毛，可以在水中游动。细胞表层的分子可以检测水中的养分。因为池塘不同区域的养分含量不同，因此从细胞的一端到另一端的养分含量值存在梯度渐变。当

细胞在池塘中游动时，它能检测到养分含量的变化。这是单细胞生物的世界中有关结构的一种简单形式。细胞能够利用其化学敏感性，游到富含养分的区域。我们可以说，这个简单的生命体正在作一种预测：预测往哪边游才能找到更多的养分。这种预测是否有记忆的参与呢？有。记忆就存在于生命体的DNA中。在单细胞动物的生命中，并不需要学习如何利用养分的渐变。确切地说，这个学习发生在演化过程中，并存储于DNA里。如果世界的结构突然发生变化，这种单细胞动物是无法学习适应的。它无法改变自己的DNA以及由其决定的行为方式。对于这类物种来说，学习只能发生在演化过程中，通过很多代来实现。

那么单细胞动物有智能吗？如果以人类智能作为标准，答案为否。但是，单细胞动物无疑属于最早的能够利用记忆和预测而更成功地繁衍的物种。因此从更学术的角度来看，答案为是。问题并不在于将有些动物标注为有智能，而将其他动物标为没有智能。所有生物都在使用记忆和预测，只是具体方法和复杂程度不同而已。

植物也会通过记忆和预测来利用世界的结构。一棵树在生根发芽时就开始了预测。它会根据先辈的经验预测在哪里能够找到水源和矿物质。当然，树是不会思考的，它的行为都是自动的。但这些物种也采用了与单细胞生物相同的方式来利用这个世界的结构。每种植物会采用不同的行为来利用这个世界结构中的稍有不同的部分。

最终，植物进化出了基于缓慢释放的化学信号的通信系统。当昆虫啃食一棵树的时候，树会通过维管系统释放化学物质，传递给这棵树的其他部分，触发防御系统，例如产生毒素等。通过这样的通信系统，树木可以产生相对复杂的行为。神经元也许就是为了能够比植物的维管系统更快通信而进化出来的。你可以把神经元看作一个拥有自己的维管附器的细胞，但它并不通过这些附器缓慢地传输化学物质，而是开始采用电化学的动作电位来传递信息，传输速度更快。最初，快速突触传递和简单神经系统并没有学习功能。它们只是为了更快地传递信号。

伴随着演化的进程，有趣的事情发生了。神经元之间的连接变得可以修改。一个神经元能够根据最近发生的事情，选择发送还是不发送信号。生命体可以在其生命历程中改变行为模式。神经系统和行为模式具备了可塑性。由于能够迅速构建记忆，动物可以在生命历程中学习这个世界的结构。这样的话，如果世界突然发生变化，例如出现新的捕食者，动物们就不必固守基因所确定的行为模式，因为那些行为说不定已经不合时宜了。可塑的神经系统成为生物进化的巨大优势，从而引发了新物种的爆发，包括鱼、蜗牛和人。

如第三章所说，所有的动物都有旧脑，在旧脑上面就覆盖着大脑皮层。大脑皮层是最晚进化出来的神经组织。但通过大脑皮层的层次结构、恒定表征以及类比预测，哺乳动物比其他没有大脑皮层的动物能更好地利用这个世界的丰富结构。我们的祖先在大脑皮层的帮助下能够织网捕鱼。而鱼类却无法学习到被渔网网住就意味着死亡，也不能制作工具破网。所有的哺乳动物，从老鼠到猫再到人，都有大脑皮层。它们都有智能，只是程度不同。

人类智能有何不同？

记忆-预测框架为这个问题提供了两个答案。第一个答案很直观：我们的大脑皮层比较大，例如比猴子或狗的都大。人类大脑皮层差不多有一块大的餐巾那么大，因此可以学习更复杂的世界模型，并作出更复杂的预测。与其他哺乳动物相比，我们可以进行更深层次的类比，看到结构之上的更多结构。假如我们找对象，我们不仅会看像健康情况这样的简单属性，还会了解他们的朋友和父母，观察他们开车和谈吐，判断他们是否诚实。我们会考察这些次要属性，来尝试预测我们潜在的配偶未来表现如何。股市交易员会寻找交易模式中的结构。数学家会寻找数字和方程中的结构。天文学家会寻找恒星行星运行中的结构。我们拥有更大的大脑皮层，这能让我们知道，我们的家庭是城市的一部分，城市

是地区的一部分，地区是地球的一部分，而地球则是宇宙的一部分。这就是结构背后的结构。其他哺乳动物都无法作如此深层的思考。我相信我的猫肯定对我家之外的世界没有任何概念。

人类与其他哺乳动物在智能上的第二点不同在于，我们拥有语言能力。有很多书都在介绍所谓语言的独特性质，以及语言是如何发展而来的。但是实际上，语言可以很好地契合记忆-预测框架，不需要任何特殊的配置或专门的语言机制。我们所讲和所写的单词都不过是这个世界中的模式而已，正如旋律、汽车和房子一样。语言的句法和语义也都与其他对象的层级结构类似，并无二致。正如我们会将火车的声音和火车的图像联系起来那样，我们也会将所说的单词与我们记忆中与之对应的物理或语义对象关联起来。通过语言，我们可以将记忆传送到其他人脑中。语言是一种纯粹的类比，通过语言可以让其他人体验和学习他们并未真正见过的事情。语言的形成需要大型的大脑皮层，这样才能够处理句法和语义的嵌套结构。语言的形成还需要更加充分发展的运动皮层和肌肉系统，这样才能让我们发出更清晰的声音，或者做出更细致的手势。通过语言，我们可以将一生中所学到的模式传给我们的孩子。语言，无论是通过说、写还是其他文化传统的呈现方式，已经成为我们世代传递关于这个世界的知识的方式。如今，印刷或者电子通信能让我们把知识分享给遍布世界的数百万人。而没有语言的动物根本无法向他们后代传递如此多的信息。一只老鼠会在一生中学到很多模式，但它无法将这些新信息详细地传给后代，它不可能说："嘿，乖孩子，看这就是我老爸教我怎么避免电击的。"

因此，我们可以把智能划分为3个时期，每个时期都使用了记忆和预测。第一个时期的物种使用DNA作为记忆媒介。这些个体无法在生命中进行学习和适应。它们只能通过DNA向后代传递关于这个世界的记忆。

第二个时期的物种开始能够修改神经系统，从而能快速形成记忆。

这些物种能够学习世界的结构，并在生命中不断调整行为来适应世界。但是，它们仍然无法通过直接观察之外的方式将这些知识传递给后代。大脑皮层的出现和扩展就是出现在这个时期，但尚未得到明确的定义。

第三个时期，也是最后一个时期，那就是人类了。这开始于语言的出现以及大型大脑皮层的扩展。我们人类可以在生命中学习很多关于世界的结构，并能用语言有效地与其他人交流。你和我现在就是在进行这种交流。我耗费了一生中的很长一段时间来探索大脑结构，以及这种结构是如何带来思想和智能的。通过这本书，我将我学到的知识传递给了你。当然，如果不是学习了成百上千的科学家们积累的知识，我也无法实现这一点，而这些科学家也曾向别人学习过，这种学习过程纵贯古今。我可以学习和吸收其他人的所见所思，并在此基础上添加我的思考和观察。

我们是地球上最具适应能力的生物，只有我们才有能力广泛传播关于这个世界的认识。正是由于我们能够充分学习和利用世界的结构并互相交流，人类数目才急剧膨胀。我们可以在任何地方繁衍生息，不论是雨林、沙漠、冻土，还是混凝土建筑林立的城市。大型大脑皮层和语言的结合造就了人类的成功。

什么是创造力？

我经常被问到与创造力有关的问题。我怀疑这可能是因为很多人认为创造力是机器无法做到的事情，因此创造力是对建造智能机器这整个想法的挑战。那么什么是创造力呢？本书已经回答过好几次了。创造力并不是出现在大脑皮层某个特定区域的什么东西。它也不像情绪或平衡感那样，是由大脑皮层外的特殊结构和回路控制的。创造力是每个大脑皮层区域的内生属性。它是预测的必要组成部分。

怎么可能是这样呢？难道创造力不应该是某种非凡的品质，需要更高的智能和天赋才能实现吗？实际上并不需要。创造力可以简单地定义为类比预测，这在你的大脑皮层中随时随处都在发生着。创造力是个连续值。它包括从每天发生在感觉区的简单感知行为（例如，在新的音调上听一首歌），到发生在大脑皮层最顶端的困难而罕见的天才行为（例如，以全新的方式创作交响乐）。在基本层面上，日常的感知行为与罕见的才华横溢是相似的。只是因为日常行为太常见了，所以我们注意不到它们而已。

现在，你已经对下面3个方面有了基本了解，包括我们如何构建恒定记忆，如何使用恒定记忆作出预测，以及如何对我们过去从未经历的事情作出预测。也请记住，我们的恒定记忆也是事件序列。我们通过把恒定记忆认为接下来应该发生的事情与当前发生的细节相结合，来作出预测（回顾一下我们关于火车何时到达的比喻）。预测就是将恒定记忆序列应用于新的情境。因此所有的大脑皮层预测都是类比预测。我们通过类比过去来预测未来。

想象你在一家陌生的餐馆就餐，你想洗一下手。即使你没有进过这座建筑，你的大脑仍然会预测这家餐馆有洗手间可以洗手。你的大脑是怎么知道的呢？那是因为你去过的其他餐馆都有洗手间，所以大脑通过类比认为这家也有。而且，你还知道到哪儿去找，要找什么标志。你预测会看到一扇门或者标志，上面有标出男性或女性的符号。你预期卫生间会在餐厅的背后，要么在吧台旁边，要么在大厅对面，但一般不会在就餐区。因此，虽然你从没来过这家餐馆，但你通过类比其他餐厅，你就能找到洗手间。你不会随便乱走。你会寻找预期的模式，从而让你能快速找到洗手间。这种行为就是创新行为，它通过类比过去而预测未来。虽然我们一般不会认为这是创新，但它实际上就是。

最近我买了一款电颤琴。我有一台钢琴，但从来没玩过电颤琴。带回电颤琴的那天，我从钢琴上取来一页乐谱，放在电颤琴上，开始弹奏

简单的旋律。我能做到这点其实没什么了不起。但在基本层面上，这也是创新行为。想想其中发生了什么。电颤琴是跟钢琴很不一样的乐器。电颤琴是金色的金属条，而钢琴是黑白琴键。金属条很大，而且尺寸逐渐变化；而钢琴键很小，而且只有两种尺寸。金属条分为两排，而黑白钢琴键交错排列。在钢琴上我用手指弹奏，而在电颤琴上我要用琴槌弹奏。弹奏电颤琴和钢琴所需要的肌肉和动作都完全不同。

那么，我怎么就能在陌生乐器上弹奏音乐呢？答案在于，我的大脑皮层在电颤琴金属条和钢琴琴键之间进行了类比。这种相似性能让我弹出曲调。这跟用新的音调唱歌没有什么不同。在这两种情况下，我们都通过类比过去所学从而知道该怎么做。我知道，你会认为这两种乐器太相似了，其实你这样想正是因为我们的大脑能够自动发现这种类比关系。如果尝试让计算机来发现钢琴和电颤琴的相似性，你就会知道这有多难了。类比预测，也就是所谓的创造力，如此普遍存在，以至于我们通常不会注意到它。

但是，当记忆-预测系统在较高的抽象层发挥作用时，当系统作出不同寻常的预测时，当系统使用不同寻常的类比时，我们就坚信这是创新了。例如，大多数人会同意，一位证明了高难度数学猜想的数学家是有创造力的。但是让我们仔细看看她的大脑做了什么。我们的数学家目不转睛地盯着方程，说："我该怎么解决这个问题呢？"如果找不出明显的答案，她就会重写这个方程。将方程用另外一种方式写好后，她开始从不同的角度来观察这个相同的问题，继续目不转睛。突然她发现这个方程的某个部分看上去很眼熟。她想："啊，我认出来了。这个方程的结构和我几年前解的另一个方程很相似。"然后她通过类比作出预测。"也许我可以运用成功解决那个旧方程的技术来解决这个新方程。"于是她通过类比之前所学解决了这个问题。这就是一个创新行为。

我的父亲曾患过一种莫名其妙的血液病，医生无法确诊。那么他们

怎么制定治疗方案呢？他们所做的其中一件事情就是，观察几个月来我父亲的血液检查数据，试图发现一些模式。（我父亲打印了一份漂亮的表格，医生们可以清晰地观察这些数据。）虽然他的症状与任何已知疾病都不相符，但也有些相似性。医生们最后提出一种混合策略的治疗方案，这些策略曾治愈过其他血液病。我猜测，这种治疗方案就是医生基于对以往治疗疾病的类比而设计出来的。需要广泛接触其他非常见疾病才能识别出这些模式。

莎士比亚的比喻是创新的典范。“爱情乃是叹息所吹起的一阵烟。（Love is a smoke made with the fume of sighs.）”“用哲学的甘乳安慰你的逆运。（Adversity's sweet milk, philosophy.）”“人们的微笑之中暗藏匕首。（There's daggers in men's smiles.）”当你看到这些比喻时，会觉得它们非常形象，但要想出这些比喻却很困难，这也是莎士比亚被誉为文学天才的原因。他一定看过很多巧妙的类比，所以才能想到这些比喻。当他写出“人们的微笑之中暗藏匕首”时，他并非在讨论匕首或者微笑。匕首是对恶意的类比，而微笑是对欺骗的类比。仅仅5个英文单词中竟有两个巧妙的类比。至少我是这样理解的。诗歌的魅力在于，能够把本来看上去并不相关的词汇或概念关联起来，从而用崭新的方式解读这个世界。它们创造出意想不到的类比，能教我们认识更高层次的结构。

实际上，高度创新的艺术作品往往由于出人意料而受到欢迎。当一部电影突破了广为熟知的角色塑造、故事情节或摄影（包括电影特效）时，你会因为它的与众不同而喜欢上它。绘画、音乐、诗歌、小说等所有的创新艺术形式，都在为打破传统、突破预期而奋斗。一件伟大的艺术作品诞生于矛盾之中。我们既希望它比较熟悉，但同时又希望它独一无二、出人意料。然而，过于熟悉就如同旧式翻新，粗制滥造；而过于独特又会让人不安，难以欣赏。最好的作品往往会打破一些常规模式，但同时会给我们带来新的模式。想象一部伟大的古典音乐，它的诉求很简单，有好的节奏、简单的旋律和乐句，每个人都能理解和欣赏它。然而，它又有些与众不同和出人意表。但是，你欣赏的次数越多，你对其

中出人意料的部分的感受就越多，例如其中反复吟咏的和声，或者音调的变换。伟大的文学作品或者电影也是类似的。你欣赏的次数越多，你所发现的创新细节和复杂结构就越多。

你大概会有这样的经历，当你看着某个东西，你会突然觉得："嗯，我好像在别的什么地方见过这个模式……"你也许曾尝试过解决一个问题，那就像你大脑中的某个恒定表征被一个全新的场景所激活。你看到了两个通常不相干的事情之间的可类比性。我也许会发现，宣扬一个科学观点就像兜售一个商业创意，而进行政治改革就像抚养孩子。如果我是个诗人，瞧，我想到一个新的比喻。如果我是名科学家或工程师，我为一个久而未决的问题提出了新的解决方案。创造力，就是将你曾经经历和一生所学的所有模式进行混合和匹配。这就像是说："这个有点像那个。"这种神经机制普遍存在于大脑皮层中。

有些人会比其他人更具创造力吗？

我经常听到与创造力相关的一个问题："如果所有的大脑都天生具有创造性，那么我们在创造力方面为什么会有差异？"记忆-预测框架给出了两个可能的解答。一个与先天属性有关，一个与后天培养有关。

在后天培养方面，每个人的人生经历都不相同。因此，每个人的大脑皮层会建立起不同的世界模型和记忆，也会作出不同的类比和预测。如果我经常听音乐，我就能在新的音调上唱歌，或者在新的乐器上演奏简单旋律。如果我从来没听过音乐，我就无法作出这些预测。如果我研究过物理，我就能通过物理定律的类比来解释日常物体的运动。如果我是和狗一起长大的，那么我就更容易发现与狗有关的类比，能更好地预测狗的行为。有些人之所以在社会环境、语言使用、数学或外交等方面更具创造力，都是因为受到他们成长环境的影响。我们的预测，乃至我们的才能，都是在我们的经历的基础上建立起来的。

在第六章我描述了记忆如何沿着大脑皮层逐级下移。你接触的某些模式越多，这些模式就会在越低层重构。这样你就能够在高层学习更高阶、更抽象的对象之间的关系。这对于形成专门知识至关重要。与非专家相比，一位专家可以通过反复实践和接触，来发现更加精细微妙的模式，如在20世纪50年代后期生产的汽车的尾翅形状，或者海鸥喙上斑点的尺寸。专家能够识别模式之上的模式。从根本上讲，我们能够学习的内容有一个物理限制，那就是我们大脑皮层的大小。但是，作为人类而言，我们的大脑皮层比其他物种都要大，我们在学习方面有着极大的自由度。这完全取决于我们在人生中会接触到什么。

在先天方面，大脑存在着物理差异。当然有些差别是由基因决定的，如皮层区域的尺寸（在V1区的总面积方面，个体之间会存在超过3倍的差异）和脑半球优势（女性的左右脑之间的连接比男性更粗）等。在不同个体中，有的大脑会有更多细胞，或者有不同的连接。阿尔伯特·爱因斯坦的创造性天才不可能只是受到了他年轻时在专利办公室工作的激励环境的影响。最近对他的大脑（人们一度认为他的大脑丢失了，但几年前被发现保存在一个罐子中）的分析表明，他的大脑明显与众不同。他的大脑中每个神经元平均拥有更多的支持细胞，又称作神经胶质细胞。人们还发现在他的大脑顶叶中有不同于常人的回间沟模式，而该区域一般认为对于数学能力和空间推理非常重要。他的大脑也比大部分人脑宽15%。我们可能永远也无法知晓为什么爱因斯坦能够这么聪明而富有创造力，但可以百分之百地肯定，这其中一定有基因的影响因素。

无论在聪明大脑和普通大脑之间有什么不同，我们都是有创新性的。通过实践和学习，我们就能够提升自身的技术和才能。

你能够训练自己变得更有创造力吗？

是的，绝对可以。我已经找到好几种方法，能够培养在面对问题时

寻找有用类比的能力。首先，你需要假设你面对的问题是有答案的。人们往往很容易就放弃。你需要坚信有个解决方案等着你来发现，而你也必须坚持长时间地思考这个问题。

其次，你需要拓展思维，让大脑畅想。你需要给你大脑以足够的时间和空间来找到解答。找到问题的解答就像找到这个世界的某个模式，或者就像找到你大脑皮层中与这个问题相类似的模式。如果你迷上了一个问题，记忆-预测模型会建议你尝试不同的方式来看待这个问题，从而让你更有可能从你的经历中找到与它最相似的类比。如果你只是坐在那儿反复盯着看，就不可能有太多想法。你可以尝试将问题的几个部分用不同的方式重新组合。当我玩纵横拼字游戏时，我会不断将字母拼贴打乱。我并不是指望这些字母能碰巧拼成我想要的单词，只是不同的字母组合能够提示我想起某些单词或单词片段，它们很有可能就是答案的一部分。如果你正在看一幅画却看不出画的是什么，你可以尝试颠倒一下，换一下颜色，或者换一个角度。例如，当我在思考V1区中的不同模式是如何实现IT区中的恒定表征的问题时，我被卡住了。因此，我将问题翻转过来，思考IT区中的恒定模式如何为V1区带来不同的预测。将问题翻转过来思考的方式立竿见影，最终让我相信V1区不应该被看作单一的皮层区域。

如果你被一个问题卡住了，试着离开一小会儿，做点别的事情，然后再回来重新开始，并用新的方式思考这个问题。如果你尝试足够多次的话，新的想法迟早会送上门来。这也许会花上几天或几周的时间，但是最终会发生的。思考的目标是从你过去和现在的经历中找到类似的场景。要想成功找到答案，你必须经常面对问题陷入沉思，但也需要做些别的事情，这样大脑皮层就有机会找到可类比的记忆。

我再举个例子，来说明如何用新的方式思考问题来找到全新解答。1994年，我的同事和我一起尝试解决掌上电脑的文本输入问题。每个人都把注意力放在手写识别输入软件上。他们说：“看，你平时是在纸上

写东西的，所以你应该也可以在计算机屏幕上写东西。”然而不幸的是，手写识别真的很难。这是人脑认为很简单而电脑并不擅长的众多事情中的又一件。其原因在于，人脑使用记忆和当前语境来识别手写的内容。手写的单词和字母本身很难识别，但在具体语境下就变得很容易了。而计算机所使用的模式匹配的方法则无法胜任这个任务。我设计了几款采用传统手写识别技术的计算机，都做得不够好。

为了提高手写识别软件的性能，我奋战了多年时间，但一直被卡在那里。直到有一天，我冷静下来，尝试从另一个角度来思考这个问题。我试着寻找与之类似的问题。我对自己说：“我们在台式计算机上是如何输入文本的呢？我们是通过键盘输入的。那么，我们是怎么学会键盘输入的呢？实际上学会键盘输入并不容易。键盘才刚刚发明出来，人们要花很长时间才能学会使用键盘。在打字机式的键盘上学习盲打很难，也不直观，这跟手写完全不同，但是有几百万的用户都在学习键盘输入。为什么呢？因为它行得通。”我的思考继续作出类比：“也许我应该设计一个并不那么直观的文本输入系统，你必须通过学习才能会用，但是只要它行得通，人们就会使用它。”

这差不多就是我所走过的历程。我用键盘输入作为类比，来尝试找出用触笔在屏幕上输入文本的方法。我发现人们很乐意学习一个困难的任务（输入方式），因为它能够快速而准确地把文本输入到机器中。因此，如果我们可以提出一种用触笔输入文本的快捷可靠的新方法，即使需要学习，人们也会乐意使用它。所以我设计了一种手写字母表，能够将你的手写内容可靠地转换成计算机文本，我们将其命名为涂鸦（Graffiti）。对于传统的手写识别系统，当计算机出错时，你连原因出在哪里都无从知晓。而涂鸦系统基本不会出错，除非是你自己写错了。我们的大脑厌恶不可预测性，这也是人们讨厌传统手写识别系统的原因。

很多人认为涂鸦系统是个愚蠢至极的想法，因为它完全违背了他们

信奉的计算机理所应当的工作方式。当时人们普遍认为计算机应当去适应用户，而不是反过来。但是我坚信，人们会接受这种与键盘输入类似的输入文本的新方式。事实证明涂鸦系统是个好的解决方案，广为用户使用。到现在我还能听到人们说计算机应当适应用户，这话不总是对的。我们的大脑喜欢一致的和可预测的系统，而我们喜欢学习新的技能。

创造力会让我误入歧途吗？我能欺骗自己吗？

错误的类比总是危险的。科学史上到处都是这样的例子：漂亮的类比最后被证明是错误的。例如，著名的天文学家约翰尼斯·开普勒（Johannes Kepler）相信6个已知行星的轨道构成一个柏拉图多面体（Platonic solids）。柏拉图多面体是唯一的能由正多边形构造出的三维立体造型。有5种柏拉图多面体：四面体（四个等边三角形），六面体（六个正方形，又名立方体），八面体（八个等边三角形），十二面体（十二个正五边形）和二十面体（二十个等边三角形）。它们是由迷恋数学与宇宙之间关系的古希腊人发现的。

与所有的文艺复兴时期的学者一样，开普勒受到古希腊思想很大的影响。于他而言，有5种柏拉图多面体和有6个行星不可能只是个巧合。他在《宇宙之谜》（*The Cosmic Mystery*，1596）中说：“这个动态世界是由这些平面体表征的。总共有5个多面体，当把它们看作边界时，这5个多面体决定了6个不同的事物，因此共有6颗行星围绕着太阳运转。这也是只有6颗行星的原因。”可以看到，他作了一个很漂亮但是完全错误的类比。

开普勒继续用以太阳为中心的嵌套柏拉图多面体来解释行星的轨道。他用水星轨道定义的球体作为基准，外接一个八面体。八面体的顶端又形成一个更大的球体，形成金星的轨道。在金星的轨道外接一个二

十面体，它的顶端形成了地球的轨道。继续向外推进，地球的轨道又外接一个十二面体，顶端形成火星的轨道。火星的轨道外接一个四面体，顶端形成木星的轨道。木星的轨道外接一个六面体，顶端形成土星的轨道。这是多么地优雅和美丽。由于开普勒时代的天文观测数据准确性有限，所以他可以说服自己这个方案是合理的！（多年之后，在开普勒从已故同事第谷·布拉赫（Tycho Brahe）那里获得了高精度观测数据后，他发现自己错了，数据表明行星的轨道是椭圆形的而非圆形。）

开普勒的兴奋对科学家们来说是一个警示，其实对所有的思考者来说也是一个警示。大脑是建立模型并作出创新性预测的器官，但它的模型和预测既有可能是对的，也有可能似是而非。我们的大脑总是在观察模式和作出类比。如果找不到正确的关联，大脑也很乐意接受错误的关联。伪科学、偏执、迷信和不宽容往往源于错误的类比。

什么是意识？

这是让神经科学家们恐惧的问题之一，而在我看来其实没有必要害怕。有些科学家，例如克里斯托夫·科克（Christof Koch），就很乐意研究意识问题，但很多人认为这是一个近乎伪科学的哲学问题。我认为，哪怕只是因为很多人都对意识问题充满好奇，它也值得我们思考。虽然我无法给出完全让人满意的答案，但是我认为记忆和预测能够部分地回答这个问题。首先，作为调剂，这里先讲讲某次讨论意识问题的对话。

不久前我去长岛湾参加一个学术会议。傍晚，我们几个人拿着红酒杯来到码头，坐在水边闲谈。过了一会儿，讨论开始转向意识问题。如我所说，神经学家通常不会谈论这个问题，但是我们正处在一个美丽的环境中，红酒下肚，这个话题就被提出来了。

一位英国科学家大谈她对意识的看法：“当然，我们永远无法理解

意识。”我不同意：“意识没什么大不了的。我认为有大脑皮层就会有意识。”所有人一片寂静，紧接着争论开始了，几个科学家试图指正我的明显错误：“你不可能不觉得这个世界动人而美丽。你怎么能否认你有感知这个世界的意识呢？你必须承认你能感到某些特别之处。”为了表明观点，我说：“我不知道你们在说些什么。既然你们如此看待意识，那我不得不说我跟你们不同。我感受不到你所感受的，所以大概我不是有意识的人吧，我一定是个僵尸。”当哲学家们讨论意识的时候经常提到僵尸。僵尸被认为在物理上与人类相同，但没有意识。僵尸只是能走路和呼吸的人肉机器，而没有意识在其中。

那个英国科学家看着我说：“你当然是有意识的。”

“不，我不觉得。在你看来我可能有意识，但是我实际上没有意识。不要着急，我没事。”

她说：“好吧，难道你无法感受这个奇迹吗？”她在波光粼粼的水中划动手臂，太阳开始落山，天空现出橙红色的晚霞。

“是啊，我看到了。然后呢？”

“那你怎么解释你的主观感觉呢？”

我回答说：“是啊，我知道我在这里呢。我记忆中有很多这样的傍晚。但我没有感到有什么特别之处。所以你觉得很特别的东西，对我而言可能毫无感觉。”我试图让她不再认为意识是不可思议或不可解释的。我试图让她定义意识。

我们你来我往地一直争论到回去吃饭。我想，我没有改变任何人关于意识的存在和意义的看法。但我试图让他们明白，大部分人把意识当成了加到人脑中的某种神奇酱汁。你有个由细胞组成的大脑，倒上意识这种神奇的酱汁后，就成为了人。根据这种观点，意识是一种游离于大

脑之外的神秘之物。因此，僵尸虽然有大脑但没有意识。他们有人脑的全部配置，包括神经元和突触，但他们没有意识这种神奇酱汁。他们能做人类可以做的每件事情。从外表看我们无法区分僵尸和人类。

这种对于意识的观点是过去人们对生命力的信仰的分支，人们曾认为是生命力驱动着生命体。人们相信生命力代表了石头和植物、或者金属与少女之间的差异。现在已经很少有人相信这种观点了。现在我们已经充分理解了有机物和无机物之间的差异，所以知道并没有什么特殊酱汁。我们已经知道很多关于DNA、蛋白质折叠、基因转录和新陈代谢的知识。虽然我们并不完全了解生命系统的全部机制，但我们知道了足够多的生物学知识，从而可以远离魔法般的幻想。类似地，人们也不再认为是魔法或心灵驱动肌肉运动的。我们有长分子互相交错的折叠的蛋白质，你能够从中解码全部的生命机制。

然而，很多人却坚信意识与众不同，认为意识是不能用还原论的生物学术语解释的。我并没有专门学习过关于“意识”的专业知识，也不曾读过所有哲学家的观点。但是对于在关于意识的这场争论中人们存在的困惑，我有些自己的思考和想法。我认为有大脑皮层就会有意识。我们还可以再进一步，将意识分为两大类。第一种类似于自我意识，也是我们日常所说的“意识”。这相对容易理解。第二种是感受（qualia），就是与知觉相关联的情感，相对独立于感知输入。感受更难理解一些。

当大部分人提到“意识”的时候，他们一般说的是第一种意识。“当你从我身边走过的时候没有跟我打招呼，你意识到了吗？”“当你昨晚睡觉掉下床的时候，你意识到了吗？”“当你睡觉的时候你是没有意识的。”有些人说这类意识就是“觉知”（awareness）的意思。这两者的确很相近，但我并不认为“觉知”能涵盖这类意识的含义。我认为这类意识类似于陈述性记忆。陈述性记忆是你可以回忆起来并向他人描述的记忆。你能够用语言描述这些记忆。如果你问我上周去哪儿了，我能够告诉你。这就是陈述性记忆。如果你问我如何保持自行车的平衡，我只能

让你扶好车把、蹬脚踏板，但我没法确切地解释具体该怎么做。保持自行车平衡很大程度上与旧脑的神经活动有关，所以这不是陈述性记忆。

下面是一个思想实验，可以向你表明我们日常所说的“意识”就是陈述性记忆。我们知道，所有的记忆都存在于突触的物理变化及其连接的神经元。因此，如果我有办法逆转这些物理变化，你的记忆就会被擦除。现在假设我有个开关，可以让你的大脑回到过去的某个时间点，可以是1个小时前，24个小时前，随便多久。现在我打开机器开关，你的突触和神经元都回到了过去的某个状态。这样，我就擦除了你从那时到现在的所有记忆。

现在假设你经过了今天，在明天醒来。当你醒来时，我打开开关，擦除了过去24小时的记忆。你对前一天的记忆为零。在你的大脑看来，昨天从未发生。我告诉你现在是周三，你会反对说：“不对，这是周二啊。我非常确定。日历肯定被篡改了。不可能，这就是周二。你为什么要跟我开这个玩笑？”但是每个在周二见过你的人都会说，你在那一天是有意识的。他们看到了你，与你吃午饭，与你交谈。难道你不记得了？你会说不，这些都从未发生过。最后，在你看了你吃午饭的视频后，你开始相信，虽然你一点都不记得了，这一天的确已经发生了。这就好像你在这天做了一个毫无意识的僵尸。然而，你当时是有意识的。仅仅是擦除了你的陈述性记忆，你就不再相信你有过意识了。

这个思想实验说明了陈述性记忆与我们日常所说的“意识”是等同的。在打网球的过程中和结束后，如果我问你是否有意识，你当然会说是。但如果我擦除了你过去2个小时的记忆，你就会声明你毫无意识，也不会对那段时间里的行为负责。然而无论是哪种情况，你都曾在打网球。唯一的不同在于，你是否有我提问时的那段时间的记忆。因此，意识的含义并不是绝对的，在你的记忆擦除后，意识就会发生变化。

“感受”这种意识的问题更困难。“感受”经常出现在禅宗似的问题中，例如“为什么红色是红色，而绿色是绿色？红色对我而言和对你而

言是相同的颜色吗？为什么红色会在情绪层面与某种情感有关？它会引起我的某种特殊情感，它会引发你的什么情感呢？”

我发现这种描述很难与神经生物学建立起联系。因此我尝试重新表述这些问题。对我而言，同样难以解释的等价问题是，为什么不同的感觉会有不同？为什么视觉与听觉不同，为什么听觉又与触觉不同？如果大脑皮层处处相同，如果它的功能相同，如果它只是处理各种模式，如果没有声音或光线而只有模式进入大脑，那么为什么视觉与听觉会如此的不同呢？我发现很难描述视觉与听觉的差异，但是这种差异显然是不言而喻的，我想你也是这么认为的。虽然有的轴突表征声音，而有的轴突表征光线，但在实际处理过程中，这些轴突都是完全相同的。感觉神经元的轴突并不会携带着“光线”或“声音”的标记。

人们会有一种叫作“联觉”（synesthesia）的情形，在那时，大脑在不同感觉之间的界限变得模糊，某些声音会带有颜色，或者某些纹理会带有颜色。这告诉我们，一种感觉的性质并非是一成不变的。通过某些物理改造，大脑也可以为听觉输入赋予视觉的性质。

那么应当如何解释“感受”呢？我想到两种可能性，但均尚不完全让人满意。一种解释是，虽然听觉、触觉、视觉在大脑皮层中按照相似的原则工作，但它们在进入大脑皮层前的处理方式是不同的。在进入大脑皮层之前，听觉功能依赖一组听觉的“皮层下”（subcortical）结构来处理听觉模式。躯体感觉模式也是首先经过一组专门针对触觉的“皮层下”区域。也许“感受”就像人类的情绪那样，并非只受到大脑皮层的影响。如果不同的感官分别与不同的皮层下结构绑定在一起，而这些皮层下结构还可能与情感中心有关联，这也许就能解释我们为什么会对不同的感官输入的信息有不同的感受，虽然这还无法解释为什么最初会出现这些“感受”。

我想到的另一种可能的解释就是输入的结构，也就是输入模式自身存在着差异，这能够指导你识别和体验信息的性质。听觉系统所接收的

空间-时间模式与视觉系统所接收的空间-时间模式有着本质差异。视觉系统有上百万视觉纤维，能够携带大量空间信息。而听觉系统只有3万听觉纤维，只能携带更多的时序信息。这些差异可能就与“感受”这类意识有关。

我们能确定的是，无论意识如何定义，记忆和预测都在产生意识的过程中扮演着重要角色。

与意识有关的概念还有精神与灵魂。

当我还是小孩的时候，我经常想，如果“我”活在其他国家的另一个小孩的身体里的话会怎么样，就好像“我”是独立于我的躯体而存在似的。精神与肉体互相独立的感觉很普遍，这是大脑新皮层工作方式的自然后果。你的大脑皮层在层级记忆中创建了这个世界的模型。当这个模型自己运转的时候，思想就产生了；大脑通过回忆来产生预测，这些预测又作为感官输入的信息，进而引发新的回忆，以此类推。我们大部分的冥思苦想都不受真实世界的影响，甚至与真实世界毫无关联，它们仅由我们大脑的模型产生。思考的时候我们会闭上眼睛以求安静，就是为了避免感官输入的信息的干扰。当然我们的模型最初是通过来自真实世界的感官输入而构建的，但当我们思考这个世界的时候，我们是通过这个大脑皮层中的模型而非真实世界来完成的。

对于大脑皮层而言，我们的躯体只是外部世界的一部分。记住，我们的大脑处在一个安静的黑盒之中。它只能通过感官系统纤维传来的模式来理解这个世界。从大脑的角度来看，作为一个模式处理装置，它并不知道你的躯体与剩下的世界有何不同。在躯体与世界之间没有明确界限。大脑皮层也没有能力对大脑自身建模，因为大脑内部没有感官系统。因此我们能够理解，为什么我们的思想看似独立于我们的躯体，为什么我们看起来好像有独立的精神或灵魂。大脑皮层构建了你躯体的模型，但无法构建大脑自身的模型。你的思想，存在于大脑之中，与你的躯体和这个世界都是物理相隔的。精神与躯体是独立的，但精神与大脑

不是。

我们能够通过某些心理创伤和身体疾病清楚地看到这种差别。如果有人的手臂被切断，他的大脑模型中可能仍然有这个手臂，从而出现所谓的“幻肢”，也就是说他仍能感受到这只手臂连着他的身体。而在另一方面，如果他的相应大脑皮层受到了损伤，即使现实中他的手臂还在，他却丢失了关于这个手臂的模型。在这种情况下，他就会患上异手症，会觉得很不舒服甚至无法忍受，感觉手臂不是他自己的，而是由别人控制的。甚至会有人坚持要把这个手臂锯掉。如果我们的大脑没问题而身体生病了，我们会感到健康的精神被困在了将死的躯体内，虽然实际上是健康的大脑被困在了将死的躯体中。很自然地，人们会以为精神在躯体死后会继续存在，但是大脑死的时候精神也就死了。当大脑先于躯体死去时，这个事实就显而易见了。患有阿尔茨海默氏症或严重脑损伤的人会丧失心智，即使他们的身体仍然保持健康。

什么是想象力？

从概念上讲，想象力很简单。进入每个皮层区域的模式流，要么来自你的感官，要么来自记忆层级的较低区域。每个皮层区域都在作出预测，然后传回到较低层级。如果想要想象什么，你只需要将预测的内容转回来作为输入信息即可。即使在物理世界没有做任何事情，你也可以感受到预测的后果。“如果A事件发生了，然后B事件就会发生，然后C事件就会发生”，如此等等。当我们在准备商业会议、进行国际象棋比赛、准备体育运动或者从事其他活动的时候，我们都在进行想象。

在下国际象棋时，你会考虑如果将“马”移动到某个位置，然后看棋盘会变成什么样子。接下来你会预测你的对手怎么走，而那之后棋盘又会变成什么样子。接着你预测自己会怎么走，等等。你会在大脑中演练这些棋步及其后果。你会基于想象中的事件序列判断最初走的那步棋的

优劣，并最终作出决定。某些运动员，例如速降滑雪者，能够通过在脑中反复演练来提高比赛成绩。闭上眼睛，他们想象每个转弯，每个障碍，甚至最终胜利到达终点的样子，通过想象他们提高了成功的可能性。想象不过是规划的另一种说法。这是由我们大脑皮层的预测能力实现的。它能让我们在真正行动之前就能知道这些行动的后果。

“想象”需要大脑拥有将预测转换为输入信息的神经机制。在第六章我提到过，第6层的细胞是精确预测出现的地方。这层细胞会将预测投射回皮层的较低层级，但同时也会投射回第4层的输入细胞。因此，一个区域的输出变成它自身的输入。如前所述，长期从事大脑皮层模型研究的斯蒂芬.罗斯伯格将这个想象的回路命名为“折叠反馈”。如果你闭上眼睛想象一只海马，你皮层中的视觉区域就会激活，就像你真的看到一只海马似的。你看到的其实是你的想象。

什么是现实？

人们会既惊且忧地问：“你是说我们的大脑会创造一个世界模型？这个模型会比现实更重要吗？”

“是的，从某种意义上来说是这样的。”我说。

“这个世界难道不是存在于我的大脑之外吗？”

这是当然的。人们是真实的，树木是真实的，我的猫是真实的，你所处的社交环境也是真实的。但你对世界的理解，以及你对它的反应，都来自基于内部模型的预测。在任何时间，你都只能直接感知这个世界的很小一部分。这个微小部分能够指示你调用哪些记忆，而这个微小部分自身是不足以支持构建你当前的所有知觉的。例如现在我正在办公室打字，突然听到有人敲大门。我知道是我母亲来了，虽然我并没有真正看到她或听到她说话，但我能想象她就在楼下。在我的感官输入信息中

没有任何关于我母亲的信息，是我记忆中的世界模型通过类比过去的经验预测了她的到来。你的很多知觉中很多并不来自于感官，而是由内部记忆产生的。

所以“什么是现实”这个问题，很大程度上是我们的皮层模型能多准确地反映真实世界的问题。

我们周围世界的很多方面都是一致相容的，因此几乎每个人都有相同的内部模型。当你还是婴儿时，你就学到光线投射在圆形物体上会产生某种阴影，你也学会根据自然世界的线索判断大部分对象的形状。你会知道如果从高脚椅上扔下一个杯子，重力会导致它掉到地上。你学会了材质纹理、几何结构、各种颜色和昼夜交替的规律。每个人都会学到这些关于世界的简单物理性质。

但是，我们的世界模型还有很多是基于风俗、文化以及父母给予的教养。我们大脑中这部分模型的一致性就比较小了，甚至对不同人来说会完全不同。一个在充满父母爱心呵护的家庭中长大的孩子，长大后会认为这个世界是安全的和充满爱的。而受到父母虐待的孩子，不管以后人们对他们再好，他们都会认为这个世界是危险的和残酷的，他们不会信任别人。很多心理学都基于人们早期的生活经历、关系和教育所带来的影响，因为大脑就是从那时开始形成了这个世界的模型。

你的文化背景会彻底影响你的世界模型。例如，研究表明，亚洲人和西方人对空间和对象的知觉方式不同。亚洲人更关注对象之间的空间，而西方人更关注对象本身。这种不同导致了不同的美学诉求和不同的解决问题的方式。早期生活中接触不同的宗教信仰，会在道德、男女地位以及人生观等方面建造出完全不同的模型。很显然，这些不同的世界模型不可能都是对的，也不可能总是对的，虽然对某个个体而言它们可能是对的。好的和坏的道德理性，都有可能会被学到。

你的文化背景（和家庭经历）会让你建立起刻板印象，很不幸这是

生命中不可避免的部分。在本书中，你可以用恒定记忆（或恒定表征）来替换“刻板印象”这个词，而这丝毫不会改变原意。类比预测就是在用刻板印象作判断。负面的刻板印象会带来严重的社会后果。如果我对智能的理论是正确的，我们就无法改变人们用刻板印象思考的习惯，因为刻板印象就是大脑皮层的工作方式。刻板印象是人脑的天然属性。

消除刻板印象所带来的危害的方法是，让我们的孩子认识到错误的刻板印象，学会换位思考，具备怀疑精神。除了灌输我们所知的最好的价值观外，我们还要提升批判性思维的能力。怀疑精神，是科学方法的核心，也是区分事实与虚构的唯一方法。

* * *

到目前为止，我希望我已经说服了你，精神不过是大脑功能的一个标签而已。它并不是操纵大脑细胞的东西，也不独立于大脑细胞而存在。神经元也不过就是细胞而已。并没有神秘力量来控制神经细胞的行为。基于这个事实，我们现在可以将注意力转到如何用计算机实现大脑的记忆和预测能力上来了。

第八章 智能的未来

我们很难预见到一项新技术的终极应用。正如我们在本书中所见到的，大脑通过类比过去而作出预测。因此，我们很自然地会想着要把一项新技术用在过去的技术曾经应用的事情上。我们总想着用新工具来做熟悉的事情，只是让它们变得更快、更有效或者更廉价。

这样的例子有很多。人们曾将火车称为"铁马"，将汽车称为"无马的马车"。有几十年的时间里，电话一直被当成电报那样，只用来交流重要新闻或者紧急事件。人们直到19世纪20年代才开始经常使用电话。摄影技术最初被当作肖像画的新形式。电影最初被当作舞台表演的变种，因此20世纪的大部分时间里电影院的屏幕前都还配有幕布。

然而，新技术的终极应用却经常出人意料，远非我们的想象所及。现在电话已经发展成为无线语音和数据通讯网络，能够让地球上的任何两个人无论在何处都可以通过语音、文本和图像通讯。1947年贝尔实验室发明了晶体管。虽然人们很快就明白这是个巨大的突破，但最初也不过将它用来改进旧的应用——将真空管替换为晶体管，这样就能生产出更小更可靠的收音机和计算机。这在当时看来已经很重要，很让人兴奋，但新旧之间的主要差别仅在于机器的尺寸和可靠性方面。晶体管的革命性应用直到后来才出现。人们经过了一段时间的渐进创新后才发明出了集成电路、微处理器、数字信号处理器或存储器。类似地，1970年发明的微处理器，最初是为桌面计算器设计的，可见这最初也不过是对已有技术的替代。电子计算器是为了替代机械桌面计算器。微处理器也成为螺线管的替代者，用于某些工业控制方面，如交通信号灯切换。好多年以后，微处理器的真正威力才凸现出来。过去没人能够预见到如今常见的信息技术，例如现代个人计算机、手机、互联网、全球定位系

统，等等。

出于同样的原因，我们也不可能预见到，类大脑记忆系统会带来什么样的革命性应用。这种智能机器将改进我们生活的方方面面，我对这一点信心十足。这几乎是可以肯定的。但是要准确预测若干年后的未来技术则是不可能的。你只需读一读未来学家们多年来所作的荒谬预言就会同意这一点。在19世纪50年代，有人预测2000年人类会在地下室拥有原子反应堆，在月球上度假。但是，只要我们时刻以此为鉴，我们就能够在预测智能机器的过程中大受裨益。至少，我们可以对未来作出一些宽泛而有益的预测。

有很多问题耐人寻味。我们能够制造智能机器吗？如果可以，它们会是什么样子呢？它们会像小说中的那样长得像人一样，还是会像个人计算机那样是个黑色或米色的箱子，或是其他什么样子？我们怎么使用智能机器呢？这是不是一项危险的技术，有可能伤害到我们或者威胁到我们的自由吗？智能机器有什么明显的应用，我们有什么办法能知道它们的奇妙应用吗？智能机器对我们生活的终极影响是什么？

我们能制造智能机器吗？

是的，我们能制造智能机器，但它们可能不会是你想象的样子。也许看起来明显应当如此，但我不认为我们将会制造长得像人，或者像人那样与我们互动的智能机器。

对于智能机器的流行观念主要来自电影和书籍。它们都是些人形机器人，有的很可爱，有的很邪恶，有的偶尔笨手笨脚的。它们与我们分享感受、思想和事情，并在数不清的科幻小说情节中扮演重要的角色。经过一个世纪科幻小说的训练，人们已经将机器人看作我们未来的必不可少的部分。几代人的成长中都有机器人角色的陪伴，包括来自《禁忌

星球》的罗比，来自《星球大战》的R2D2和C3PO，以及来自《星际迷航》的海军少校达塔。即使是电影《2001：太空漫游》中的HAL，虽然没有躯体，却非常人性化，它既是太空飞船的控制者，更是人类在漫长太空旅行中的伴侣。有限功能的机器人，诸如智能汽车，探索深海的自主微型潜艇，自导航吸尘器或割草机等，都非常切实可行，终有一天会变得很常见。而像指挥官达塔和C3PO这样的机器人，在未来很长一段时间内仍然不可能成为现实。其中的原因有很多。

首先，人类精神的产生，既需要大脑皮层，也需要旧脑中的情感系统，还需要复杂的人类躯体。要想成为人类，你需要所有的生物机能，而不仅仅是大脑皮层。要想让机器像人那样交流（通过图灵测试），需要机器有真人那样的经历和情感，并像人一样生活。智能机器会有大脑皮层那样的结构，会有一些感觉，而其余的都不是必需的。如果智能机器拥有了类人的躯体，也许会很有娱乐性，但除非为它加入像人那样的情感系统和经历，否则它是不会具备人类精神的。要实现这一点非常困难，而且也不是我的目标。

其次，考虑到制造和维护类人机器人所必须付出的巨大成本，我们很难看到类人机器人的实用价值。与人类助手相比，机器人管家会更加昂贵，而且没那么有用。也许这个机器人管家显得很“智能”，但不可能像作为人类一员的人类助手那样交流融洽和善解人意。

蒸汽机和数字计算机曾经诱发了关于机器视觉的设想，但没有取得什么成果。同样，当我们考虑制造智能机器时，很多人觉得自然应该制造类人机器人，但这是不可能的。机器人这个概念诞生于工业革命时代，然后不断被小说完善。在发展真正的智能机器方面，我们不应该向它们寻求灵感。

如果智能机器不会走路也不会说话，那么它们应该长什么样呢？生物进化过程表明，如果我们的感官附加上了层次记忆系统，记忆就能够对世界建模并预测未来。师法自然，我们应该沿着相同的路线制造智能

机器。以下是构建智能机器的方案。我们首先要用一组感官提取来自这个世界的模式。我们的智能机器可能拥有与人类不同的感官，这些感官甚至可能“存在”于跟我们不同的世界中（后面将详细讨论）。所以不要想当然地认为它必须有眼睛和耳朵。接下来，将这些感官与层次记忆系统相连，该记忆系统采用与大脑皮质相同的工作原理。然后我们就像教孩子那样训练记忆系统。通过反复训练，我们的智能机器将按照它的感官所观察到的世界建立它的世界模型。这个过程不需要也没有机会让人手工输入世界规则、数据库、事实或者任何高层概念，而这些正是人工智能的祸根所在。智能机器必须通过观察它的世界来学习，在必要的时候也可以包括来自老师的输入信息。一旦我们的智能机器创建了它的世界模型，它就能根据过去的经验作出类比，对未来的事件作出预测，对新的问题提出解决方案，并向我们提供这些知识。

从物理上讲，我们的智能机器可以内置于飞机或汽车，或者就放在计算机房的机架上。与人类不同，人类大脑必须在人的身体内，而智能机器的存储系统却可以远离感官系统（也可以远离“躯体”，如果它有的话）。例如，一个智能安全系统的传感器可以遍布工厂或小镇，但与传感器连接的层次记忆系统可以被放置在一栋楼的地下室内。因此，一台智能机器可以有很多种不同的物理实现形式。

没有理由让智能机器一定要有人类那样的长相、行动或感官。它的智能体现在，它能够通过层次记忆模型理解它的世界，与之互动，并能够像你我这样思考关于世界的事情。正如我们将要看到的，它的思想和行动可能与人类完全不同，但它仍然是有智能的。“智能”的衡量标准是层次记忆的预测能力，而不是类人行为的相像程度。

* * *

让我们将注意力集中到制造智能机器时将会面临的主要技术挑战。想要制造智能机器，我们需要构建像大脑皮层那样的大型的有层次结构的记忆系统。我们会遇到规模和连通性上的挑战。

第一个问题是规模问题。大脑皮层有32万亿个突触。假设我们用2个比特（这样每个突触有4个可能取值）表示每个突触，而每个字节有8个比特（因此一个字节能够表示4个突触）。那么我们大概需要8万亿个字节的内存。现在一台个人计算机的硬盘有1000亿个字节，所以我们需要80个硬盘才能达到大脑皮层的规模（这些都是粗略估计，不是精确数字）。不过在实验室环境下，这样大的内存是绝对可以实现的，我们不会被这个1000倍的问题所困扰，但这样的机器也就无法放进你的口袋中或者安置在面包机里了。重要的是，尽管这在10年前还的确是无法想象的事情，但现在看来，我们所需的内存规模已经不是不可能的了。而事实上我们并不需要重新创造出整个大脑皮层，对很多应用而言只需很少一部分就够了。

我们的智能机器需要大量内存。开始我们可能需要用硬盘或光盘来构建内存，但最终我们希望用硅造芯片来构建，因为硅片尺寸较小，能耗较低，而且经久耐用。对于制造出足够容量的硅片内存来构建智能机器而言，只是个时间问题。事实上，智能记忆与传统计算机内存相比有一个优势。半导体工业的经济形势是基于芯片错误比率的。对于很多芯片而言，即使一个错误也能让芯片变得毫无用处。因此我们将生产出好芯片的百分比称为收益。这决定了某个芯片设计是否能投入生产、销售并获得利润。由于出现错误的几率随着芯片尺寸的增长而增长，目前的大部分芯片比邮票还要小。工业界正在不断提高单个芯片的内存容量，并保证芯片尺寸不会变大，但这是以牺牲芯片性能为代价的。

但是智能记忆芯片将具有天然的容错能力。你的大脑也不会将重要数据只保存在某个单独的部分。大脑每天都会失去几千个神经元，但一生中你的脑力的降低速度都非常慢。智能记忆芯片也会采用与大脑皮层相同的工作方式，因此即使有部分出现异常，整个芯片仍能保持正常工作，这就有商利可图。这种像人脑一样的天然容错能力很有可能让人设计出比当前计算机内存尺寸和密度都要大得多的芯片。这样也许就能更快实现将大脑放到芯片中的愿景，比现在趋势所预期的更快。

第二个需要克服的难题是连通性问题。真正的大脑中有大量的皮层下白质。如前所述，这些白质实际上是由上百万个树突组成，它们就在薄薄的大脑皮层之下，实现大脑皮层不同区域之间的互联。一个细胞可能会与5000到10000个其他细胞相连。如果采用传统的芯片工艺技术，几乎不可能实现这种规模的互联。硅芯片是通过若干层金属镀层实现互联的，每层之间都由绝缘层隔开（这跟大脑皮层中的分层没有任何关系）。这些金属镀层实现了芯片中的互联，由于同一层中的连接互相不能交叉，所以连接总数会受到限制。这样的工艺是根本无法实现类人记忆系统的，因为后者需要上百万条连接。硅芯片无法实现脑白质的强大功能。

虽然需要大量的工程和实验，但我们知道解决这个难题的基本思路。电线传输信号的速度要比神经元树突快得多。芯片上的电线是可以共享的，因此可以用来实现多个不同的连接，而在大脑中每个树突只属于一个神经元。

电话系统就是这样的案例。如果我们在任何两个电话之间都拉一条电话线，那么整个世界就要被电话线覆盖了。实际上我们让所有电话共享较少的高容量电话线。只要每条电话线的容量远大于传输单个通话所需的容量，这个方法就能奏效。而电话系统能够满足这个要求：一条光缆可以同时传输一百万个通话。

大脑已经在细胞之间形成固定的树突，而我们可以像电话系统那样通过共享连接的方式制造智能机器。信不信由你，多年前就已经有科学家思考大脑芯片的连通性问题了。虽然现在对我们而言大脑皮层的工作机制还是个谜，但研究者们认为总有一天我们能解决这个谜题，那时我们就不得不面对连通性这个难题了。这里，我们不必介绍这些解决方案。我们可以说，连通性问题是我们制造智能系统过程中最大的技术挑战，而我们有办法解决它。

一旦解决了这些挑战性问题，就没有什么能阻碍我们制造智能系统

了。但是要想把这些系统做得更小巧、更便宜、更强大，我们还需要解决很多问题，但这些都难不倒我们。计算机从房间大小发展到可以放到口袋里花了50年的时间。而我们现在是从先进技术开始起步的，所以对智能机器而言，这种改进会更快。

我们应该制造智能机器吗？

经过21世纪后，智能机器将从科幻世界进入现实。在这之前，我们应该考虑一下伦理问题：智能机器可能引发的危险会不会比带来的益处更大？

机器能够自主思考和行动的前景曾经一度让人们焦虑。这是可以理解的。新的知识和技术在刚出现的时候总是会让人们害怕。人们有着丰富的想象力，会想象新技术可能会用很多可怕的方式控制我们的身体，让我们毫无用处，或者让我们生命毫无价值。但是历史证明，这些恐怖的场景从来没有发生过。当工业革命出现时，我们害怕电能（记得弗兰肯斯坦吗？）和蒸汽机。自身拥有能量并且可以做复杂运动的机械，看起来非常神奇但又暗藏险恶。而如今的电力和内燃机，都不再陌生或者险恶。它们就像空气和水一样，成为我们生活环境的一部分。

当信息革命到来时，我们也很害怕计算机。有无数的科幻小说在讲强大的计算机或者计算机网络自发地产生了自我意识，然后开始进攻人类。但是现在计算机已经融入我们的日常生活，这让过去的担心显得非常滑稽，因为如果真是那样的话，你家的计算机、互联网跟自动柜员机都有相同的可能性——产生自我意识。

任何技术都可能被用来做好事或者坏事。当然，有些技术更容易被滥用或者带来灾难。原子能很危险，无论是被用来制造核弹还是发电，因为一次事故或滥用都有可能杀死数百万人。而且，虽然核能很有价

值，但它是可替代的。运载工具技术可以制造坦克和战斗机，但也可以制造汽车和客机，发生一次事故或使用不当也会给很多人带来伤害。但是对于现代生活而言，汽车可能更重要一些，而且危险性低于核能。一次飞机滥用可能造成的损失要远小于核弹。也有很多技术几乎百利而无一害，电话就是一例。在绝大多数情况下，电话能让人们互相更加亲近，这点好处远胜过它的任何负面作用。电力和公共健康科学也是这种技术。在我看来，智能机器将是我们开发的危险最小、好处最多的技术之一。

当然还有一些人，例如太阳微系统公司（Sun Microsystems）联合创始人比尔·乔伊（Bill Joy），担心我们制造的智能机器人会失去控制，占领地球，并根据它们自己的意愿来改造它。这个场景让我想到动画片《魔法学徒》中的那把魔法扫帚，能够从碎片中重生，并不分日夜地工作，从而引发了灾难。类似地，有些人工智能乐观主义者所提出的延长寿命的预言也让人不安。例如，雷·库兹威尔（Ray Kurzweil）曾经谈到，纳米机器人能够在你的大脑中游走，记录每个突触和连接，然后将这些信息报告给超级计算机，计算机就可以通过重新配置把自己变成你！这样，你就有了一个“软件”版本的自己，而这个软件可以长生不死。智能机器胡作非为，将人脑上载到机器中，如此种种，这些关于机器智能的预言反复出现，层出不穷。

制造智能机器不同于制造自我复制的机器。这两者没有任何逻辑联系。无论是大脑还是计算机都无法直接实现自我复制，类人脑的记忆系统也无法做到。虽然智能机器的一大优点是我们可以大量生产它们，但这与细菌和病毒的自我复制方式完全不同。自我复制不需要智能，智能也不需要自我复制。

此外，我严重怀疑我们是否有能力将人类心灵复制到机器中去。据我所知，目前还没有办法能够记录数万亿个与“你”有关的细节。我们需要记录并重新建立你所有的神经系统和躯体，而不单是你的大脑皮层。

而且，我们还要了解它们都是如何工作的。当然，也许到未来的某一天，我们能够做到这一点，但是这面临的挑战要远远超过了解大脑皮层的工作原理。搞清楚大脑皮层算法并从头构建成智能机器是一回事，而扫描大脑的无数细节，并复制到机器中则是完全不同的另外一回事。

* * *

除了自我复制以及复制人脑之外，人们对智能机器还有其他方面的担心。智能机器会不会像原子弹那样有巨大的杀伤力？智能机器会不会变得邪恶，与人类作对，就像《黑客帝国》和《终结者》里面那样？

对这些问题的答案都为否。作为信息设备，类人记忆系统将是我们最有用的技术。但是就像汽车和计算机那样，智能机器不过是工具而已。它们拥有智能，但并不代表它们有毁坏物品和控制人类的特殊能力。正如我们不会把世界上的核武器交给一个人或计算机控制一样，我们也必须注意不要太依赖智能机器，因为它们就像其他技术一样也会出错。

这又把我们引到一个恶毒的问题上来。有人认为有智能基本上就等于拥有人的意识。他们害怕智能机器会厌恶自己被人类“奴役”，因为人类也讨厌被别人奴役。他们害怕智能机器会试图控制世界，因为历史上的聪明人都曾试图控制世界。其实，这些担心都是以一个错误的类比作为基础的，他们把大脑皮层算法与旧脑中的情感驱动（如恐惧、偏执和欲望）混为一谈。而智能机器并没有这些功能。它们不会有个人的野心。它们不会渴望财富、社会认可或者情感上的满足。它们也不会有食欲、癖好或者情绪障碍。智能机器不会有任何类似人类那样的情感，除非我们刻意设计这样的功能。在人类智力所不及的地方才是智能机器的用武之地，在这些地方，我们的感官不灵或者感到工作单调。而这些活动一般不涉及情感。

智能机器既有简单的、单一功能的系统，也有非常强大的、超人的

智能系统，但是除非我们让它们变得与人类一样，否则它们就不会。也许有一天我们不得不限制智能机器的应用领域，但要到那一天还有很长的路要走。而当那一天来临的时候，与我们现在所面临的遗传学和核技术等道德问题相比，智能系统的伦理问题也许已经变得相对容易处理了。

为什么要制造智能机器？

现在我们考虑这个问题：未来智能机器可以做什么？

我经常被邀请作报告展望移动计算的未来。会议组织者会让我描述5年或20年后掌上电脑或手机可能长什么样。他们希望听到我对未来的设想。但是我做不到。我尽量避免对未来作出预测。为了表明这一观点，我曾经戴着巫师帽，拿着水晶球走上舞台。我解释说，没有人能够准确地预测未来。预见未来的水晶球是假的，无论谁假装知道未来会发生什么都注定会失败。然而，我们却可以了解大致的发展趋势。如果你能对全局有所了解，那么无论细节如何，你都能成功地赶上潮流。

关于技术趋势最有名的例子是摩尔定律（Moore’s Law）。戈登·摩尔（Gordon Moore）准确地预测了硅芯片上可容纳的电路元件的数目会每一年半翻一番。这里，摩尔并没有明确说明这种芯片指的是内存芯片、中央处理器还是别的什么芯片。他也没有说明芯片将被用于什么产品。他没有预测芯片是装在塑料或陶瓷中，还是粘在电路板上。他也没有提及制造芯片的各个流程。他只是预测了大致的发展趋势，而且他预测对了。

我们现在无法预测智能机器的终极应用。我们无法预测任何具体的细节。如果我或其他任何人对智能机器的用途作出了详细预测，都将不可避免地在未来被证明是错误的。但我们还是可以作出一些预测，总比

无奈地耸耸肩要好。有两种思路进行预测。第一种是考虑类大脑记忆系统的短期应用，这种应用显而易见，但不太有趣，可以先试试看。第二种是考虑长期趋势，像摩尔定律那样，帮助我们想象未来可能成为生活的一部分的那些应用。

让我们先从短期应用开始。这些应用都很显而易见，就像用电子管替换收音机的真空管、用微处理器制造计算器一样。我们可以看看那些人工智能曾经尝试过但尚未解决的领域，如语音识别、视觉和智能汽车等。

* * *

如果你尝试过使用语音识别软件将文本输入计算机，你就知道这些软件的性能是多么地有限。就像赛尔的“中文小屋”一样，计算机根本无法理解这些语音。在使用过一些语音识别产品后，我变得越来越沮丧。如果房间里出现任何噪音，无论是铅笔掉在地上还是有人跟我说话，都会导致多余的单词出现在计算机屏幕上。语音识别的错误率太高了。软件识别出的单词经常毫无意义。“记住让玛丽摔倒，沼泽可以随时被激起。”[1]连小孩都知道这句话是不对的，而计算机却做不到。类似地，自然语言接口是计算机科学家多年来的研究目标。想法是让你能够用自然语言告诉计算机或其他设备你想要什么，并指示机器工作。对着个人数字助理或PDA，你可能会说：“将我女儿周日的篮球比赛挪到早上10点。”传统AI技术不可能做好这种事情。即使计算机能够识别每一个单词，如果想完成这个任务，它还需要知道你女儿在哪里上学，知道你的意思是接下来的这个周日，并且知道这场篮球比赛指的是之前日程中写的那个“门罗对阵圣乔”。或许，你可能想让计算机收听电台广播查看是否提到了某个特定产品，而电台播音员只描述了这个产品，但没有提及它的名称。你我都知道播音员提到了这个产品，但计算机却做不到。

以上这些应用以及许多其他的应用，都需要机器能够听懂口语。但

计算机无法完成这些任务，因为它们根本不明白你在说什么。它们只会把语音模式生搬硬套到单词上，根本不了解这些单词的意思是什么。试想如果你学习某门外语中的单词发音，但不知道单词的意思，然后我让你把用这个语言说的一段对话记录下来。随着对话的进行，你根本不知道在讨论什么，而只能孤立地写下每个听到的单词。然而，单词之间会发生重叠和干扰，噪音会掩盖部分声音。你会发现很难把单词区分开并识别它们。这些困难就是现在语音识别软件在努力克服的。工程师们已经发现，通过使用单词转换概率，可以在一定程度上提升软件准确率。例如，工程师采用语法规则来帮助选择同音词。这是一种非常简单的预测，但系统仍然无法实际使用。现在语音识别软件的应用场景极为受限，在任何时刻你说的单词数量都要受到严格限制。然而，人类却可以轻松地完成许多语言相关的任务，因为我们的大脑皮层不仅理解单词，理解句子，还理解说话时的语境。我们对思想、短语和单词都会作出预测。我们的大脑皮层关于这个世界的模型会自动完成这些工作。

因此，我们可以预期，类皮层记忆系统能够将不可靠的语音识别转变为强大而鲁棒的语音理解。与概率式的单词转换不同，层次记忆能够同时追踪口音、单词、短语和思想，并以此来理解你所说的话。与人类一样，这样的智能机器可以区分不同的言语活动，例如，在房间里你和朋友之间的讨论、电话交谈，以及对一本书的编辑指令。制造这样的智能机器并不容易。为了完全理解人类语言，一台机器必须体验和学习人们所做的事情。因此，也许还需要很多年才能让智能机器像你我这样真正地理解人类语言，但短期内我们可以通过使用类皮层记忆系统来改进现有语音识别系统的性能。

视觉也是传统人工智能技术无法做到而真正智能系统可以处理的一种应用。现在还没有机器能够看懂一个自然场景（例如你眼前的世界，或者相机拍摄的照片），并能描述它看到的画面。现在在某些特别受限的领域，会有一些机器视觉的成功应用，例如在电路板上通过机器视觉对齐芯片的位置，或者与数据库中的数据做面部特征匹配，但是目前计

算机仍无法识别各种对象，更无法分析一般的场景。对你而言，可以毫无困难地环视房间并找个位置坐下，但计算机做不到。试想我们观看安全摄像头的视频画面。你能分辨出拿着礼物敲门和用撬棍撬门的不同吧？当然可以。但现在的软件做不到。所以，我们需要雇人24小时地看着安全摄像头屏幕，寻找可疑事件。对于人类而言很难一直保持警觉，但智能机器却能不知疲倦地执行任务。

最后，让我们看一下交通。汽车正变得越来越精致。它们有全球定位系统来协助你从地点A开到地点B，有传感器在天黑时开灯，有加速计帮助释放安全气囊，还有近距离传感器帮助你倒车。甚至还有汽车可以在特殊的公路上和理想的环境中自动驾驶，当然这种汽车还没有进入市场。但是，要想在各种道路和交通状况下安全有效地行驶，汽车光有几个传感器和反馈控制电路是不够的。要成为一个好司机，你必须了解交通、其他司机、汽车工作方式、信号灯以及其他很多事情。当其他汽车危险驾驶时，你需要知道危险的迹象或信号。你需要注意观察其他汽车的转向信号，并预期它要变道，而如果这个信号持续了几分钟，你要意识到那个司机可能并不知道自己的信号灯被打开了，因此它可能不会变道。你需要认识到，前面远处的青烟可能意味着发生了事故，因此你需要减速。如果司机看到一只球从车前滚过，他会自然地想到可能会有小孩跑出来追球，因此会马上把车停下来。

比方说我们现在想造一辆真正的智能汽车。我们首先会做的事情是选择一组传感器，能够让智能汽车感知到这个世界。我们不妨用摄像头收集视觉信息，在车前车后配置若干摄像头，然后用麦克风收集听觉信息，我们可能还需要添置雷达和超声波传感器，在明暗条件下均能准确地定位其他物体的大小和速度。关键是，我们不必依赖或者受限于人类的感官类型。大脑皮层算法很灵活，只要我们合理地设计层次记忆系统，无论什么类型的传感器都可以用。理论上，智能汽车能比我们更好地感知世界，因为我们可以选择适用于这个任务的感官组合。这些传感器将被连接到一个大型的层次记忆系统上。汽车设计师将智能汽车暴露

在真实世界条件下，让它用与人类相同的学习方式来学习建立自己的世界模型——只不过是在驾驶这个受限领域（例如，这个智能汽车只需要了解道路，而不需要了解电梯和飞机）——从而训练出智能汽车的记忆。智能汽车的记忆系统会学习交通和道路的层次结构，这样它就可以了解和预测它的世界中可能发生的事情，例如移动的汽车、公路标示、路障以及十字路口，等等。智能汽车的工程师们设计出的这个记忆系统，可以真正地驾驶汽车，也可以只监视你开车时会发生什么。它可以提供意见，或在极端情况下帮助驾驶，就像一个你不会反感的副驾驶员那样。一旦智能汽车的记忆系统得到全面训练，能够理解和处理任何可能发生的事情，工程师可以有两种选择：或者永久地固化记忆，这样的话，走下装配线的汽车都有相同的行为方式；或者汽车记忆在售出后仍能继续学习。作为计算机而非人脑，它的记忆可以在条件允许的时候被升级为新的版本。

我并不是说我们非要制造能够理解语言和视觉的智能汽车或机器不可。但这些都是我们可以研究和发展的不错方向，而且看起来也很有可能实现。

* * *

从我个人来讲，我对智能机器的这些显而易见的应用兴趣不大。对我而言，新技术真正的好处和让人兴奋之处在于，找到那些在过去看来不可思议的应用。智能机器将会以什么方式让我们惊讶不已，它们将有什么神奇的功能出现？我确信层次记忆会像晶体管和微处理器那样，以不可思议的方式改善我们的生活。但问题是它将如何做到这一点呢？有种办法可以让我们一窥智能机器的未来，那就是想象这项技术的哪些方面会有较好的可扩展性。也就是说，智能机器的哪些属性将变得越来越便宜，速度越来越快，或者越来越小。这些以指数级速率发展的部分将迅速超出我们的想象，并最有可能在未来技术的剧烈变革中发挥关键作用。

以指数级速率发展多年的技术有很多，包括硅内存芯片、硬盘、DNA测序技术以及光纤。这些快速发展的技术已经成为多种新产品和业务的基础。而软件也用不同的方式表现出较好的可扩展性。一个设计良好的程序，一旦完成，就可以毫无成本地无限复制。

与此相反，有些技术的可扩展性很差，如电池、发动机和传统机器人。尽管通过大量努力得到了稳步的改进，现在的机械臂并不比几年前的好太多。机器人技术的发展渐进而温和，并没有出现像芯片设计或软件生产那样指数级的增长曲线。在1985年需要用100万美元制造的机械臂，在今天不可能只用10块钱造出1000倍能力的来。同样，今天的电池并不比10年前的有太多改进。你大概可以说性能提升了两三倍，但没有成千上万倍的改进，这方面的进步非常缓慢。如果电池容量的增长与硬盘容量的增长速度相同，那么手机和其他电子产品就永远不需要充电了，而轻便电动汽车充一次电将足够行驶1000英里。

因此，我们有必要考虑类人脑记忆系统在哪些方面会出现较大扩展，能够超过我们的生物大脑。这些属性将启示该技术最终落地的应用。我认为有四个属性将超出我们人类自身的能力，它们分别是速度、规模、可复制性和感官系统。

速度

神经元的速度在毫秒量级，而硅的工作速度可以达到纳秒量级（并且仍在变得越来越快）。这是百万倍的差异，或者说是6个数量级的差异。人类大脑和以硅为基础的智能系统之间的速度差异将产生巨大影响。智能机器的思考速度将比人脑快百万倍。这样的系统将在几分钟内就能读完整个图书馆的书或者研究完海量复杂的数据，而这些任务你我可能需要好几年才能完成。这其中没有任何魔法。生物大脑在进化中有两个与时间有关的约束。一个是细胞工作的速度，而另一个是世界变化

的速度。如果周围的世界本质上是缓慢变化的，生物大脑就没有必要让思考速度快100万倍。但是大脑皮层的算法本身并不一定非要这么慢。如果智能机器需要与人类进行交谈或互动，那么它不得不放慢工作速度到人类的速度上。如果它翻动书页来阅读一本书，那么它的阅读速度会受到内容读取速度的限制。但是，当它与电子世界交流时，它就可以发挥出更快的速度。两个智能机器的交流速度可以比两个人之间快100万倍。想象一下，智能机器解决数学或科学问题的速度能比人类快100万倍。它10秒内在一个问题上想到的东西，需要你花一个月。从不感到疲劳，从不觉得厌倦，如闪电般的速度，这样的大脑肯定会大有作为，只是我们现在无法想象得到。

规模

尽管人类大脑皮层的存储规模令人印象深刻，但是智能机器的规模仍能大幅度超过它。我们大脑的尺寸受制于多种生物因素，包括婴儿头骨与产妇骨盆的直径比、大脑运转的高代谢消耗（你的大脑只有你体重的2%左右，但会消耗你呼吸氧气的20%左右）以及神经元的缓慢速度。然而，我们却能建立任意大小的智能记忆系统。而且不同于盲目、曲折的进化过程，我们会在智能系统的设计过程中深思熟虑，目标明确。人类大脑皮层的容量在未来几十年中不会有明显变化。

当制造智能机器时，我们有很多种方式增加它们的记忆规模。增加层次结构的深度可以带来更深层的理解，能够看到更高阶的模式。扩大皮层区域内的规模将让机器记住更多的细节，或者能提升感官的敏锐度，盲人就是以这种方式拥有更敏锐的触觉或听觉的。而增加新的感官和感官层次结构能够让机器构建更好的世界模型，我稍后会作更详细的讨论。

智能记忆系统的规模有上限吗？上限会是多大？这是个有趣的问

题。可以想象，当系统接近某个理论上限时，它可能会变得过于复杂而无法使用，或者开始出错。也许人脑已经接近其理论上限，但我认为这不太可能。在进化的较后期人脑才开始变得比较大，也没有迹象表明人脑已经处于稳定的最大值。无论智能记忆系统的规模峰值最终是多少，人脑都几乎不可能达到，甚至连接近这个值都不太可能。

考察这些智能系统潜在应用的一种方式是看已知的人类性能的极限。爱因斯坦无疑绝顶聪明，但他的大脑仍然是一个人类大脑。我们可以假设他过人的智慧，主要是他的大脑与典型人脑之间的物理差异的产物。爱因斯坦这样聪明的人如此罕见的原因在于，人类基因一般不能产生这样的大脑。然而，当我们基于硅芯片设计大脑时，我们可以用我们想要的任何方式来制造智能系统。它们可以拥有爱因斯坦那样的高层次思想，甚至更加聪明。而在另一个极端，白痴天才（savants）可以让我们看到智能的其他表现方面。白痴天才的智力迟钝，却有很多非凡的能力，例如近似照相机般的记忆，或者闪电般快速计算数学难题的能力。他们的大脑，虽然不是典型人脑，但仍然是人脑，采用了相同的大脑皮层算法。如果一个非典型大脑可以拥有惊人的记忆能力，那么理论上我们可以在人造大脑中添加这些功能。人类心智能力的这两个极端，不仅告诉我们在智能系统中有哪些功能应当被重建，而且也代表了我们可以超越人类最好表现的方向。

可复制性

每个新的大脑都需要从头开始成长和训练，这需要花费人类几十年的时间。每个人都必须自己学会协调肢体和肌肉群，掌握平衡和移动的基本能力，学习众多物体、动物和其他人的普遍属性，学习事物名称和语言结构，学习家庭和社会的规则。一旦掌握了这些基础知识，就会开始为期多年的正规学校教育。即使其他很多人都已经经历过这个学习的过程，每个人生命中也都还是需要经过相同的学习曲线的洗礼，而这一

切都是为了建立大脑皮层中的世界模型。

智能机器则不需要经过这么漫长的学习曲线，因为芯片和其他存储介质可以无限复制，轻松转移内容。在这个意义上讲，智能机器可以像软件一样复制。一旦某个原型系统得到圆满的训练，我们就可以任意地复制它。我们可能需要花很多年来进行芯片设计、硬件配置、训练和试错，不断完善智能汽车的记忆系统，而一旦得到最终产品，就可以进行批量生产。正如我前面提到的，我们可以选择是否允许每个复制品继续学习。对于有些应用，我们会希望智能机器按照测试过的已知方法来运行。当智能汽车知道了我们需要它知道的一切，我们就不希望它养成坏习惯，或者开始相信一些错误的类比。如果不出意外，我们希望所有相似工艺的汽车拥有相似的行为。但是对于其他应用，我们会希望类人脑记忆系统有完全的能力不断学习。例如，一个旨在探索数学证明的智能机器，需要有能力从经验中学习，能够将过去的见识应用到新问题上，具备灵活和开放性的能力。

正如我们可以共享软件的组件那样，我们也应该可以共享学习的组件。某种特殊设计的智能机器，应该可以被重新编程，用一套新的连接实现不同的行为，就如同我能下载一套新的连接装入你的大脑，让你瞬间从英语使用者变成法语使用者，或者从政治学教授变成音乐学家。人们可以与他人的工作进行交换并构建自己的系统。比方说，我已经开发并训练了一台拥有超强视觉系统的机器，而另一个人开发并训练了一台拥有超强听觉系统的机器。通过这种良好的设计，我们可以充分结合这两个系统的优点，而不必再重新进行自下而上的训练。这种共享专业知识的方式，对人类而言是根本不可能的。制造智能机器的商业模式可以遵循计算机工业的发展路线，不同领域的集团负责训练智能机器的不同种类的专门知识和能力，然后组合成各种记忆配置进行销售和交易。对智能系统进行重新编程与运行新的视频游戏或安装新的软件不会有太多不同。

感官系统

人类只有少数的几种感官。这些感官都深深植根于我们的基因，我们的身体，以及我们大脑皮层下的连接中。我们无法改变它们。有时，我们利用技术来增强我们的感官，例如夜视镜、雷达或哈勃太空望远镜。这些高科技仪器利用高超的手段进行数据转换，而本身并没有提供新的感官模式。它们将人类无法感知的信息转化为我们可以理解的视觉或听觉信息。正是由于我们大脑惊人的灵活性，我们才能够观察雷达屏幕并理解它的意义。许多动物物种拥有真正的不同感官，如蝙蝠和海豚的回声定位能力，蜜蜂感受偏振光和紫外线的能力，以及一些鱼类对电场的感觉。

我们的智能机器可以通过任意的感官方式感知这个世界，无论是自然界中已有的感官方式，还是纯粹由人类设计出来的新的感官方式。声呐、雷达和红外视觉都是非人类感官方式的典型例子，我们也许会让智能机器拥有这些感官。但这些仅仅是个开始。

更有趣的是让智能机器通过全新的方式感知世界。正如我们所看到的，大脑皮层算法从根本上只关注这个世界的模式。它不关心这些模式的物理来源。只要皮层输入不是随机的，并有足够的丰富程度或统计结构，智能系统就会形成关于这些输入的恒定表征和预测。这些输入模式没有必要与动物感官相似，甚至没有必要一定是来自真实世界的。我猜想，正是全新的感官方式，能够为智能系统带来革命性的应用。

例如，我们可以设计一个遍布全球的传感系统。试想在大陆上每隔50英里设置一个气象传感器。这些传感器就像视网膜上的细胞。在任何时间点，两个相邻气象传感器的行为将具有高度的相关性，就像视网膜上的两个相邻细胞。各种大型的气象对象，例如风暴和冷热锋，随时间不断移动和变化，就像随时间不断移动和变化的可视对象。将这些感官数据输入到大型类皮层记忆系统中，我们可以让系统学习和预测天气变

化，就像你我学会辨认可视对象并预测它们的移动那样。这个系统可以看到局部的天气模式、大型的天气模式，以及过去几十年、数年和数个小时的天气模式。假如将某些区域的传感器放置得更加密集，我们可以创建出类似视觉中央凹的区域，能够让我们的智能天气系统理解和预测微型气候。我们的天气系统会思考和了解全球的天气系统，就像你我会思考和理解物体和人那样。现在气象学家正在致力于类似的事情。他们收集不同位置的记录，使用超级计算机模拟气候并预测未来。但这种做法与智能机器的工作方式有着根本不同。他们的做法类似于计算机下棋，根本不理解任务本身的意思。而我们的智能机器则类似于人类下棋，深思熟虑而善于理解。智能机器能够发现人类自身无法发现的天气模式。直到19世纪60年代人们才发现厄尔尼诺现象。而我们的智能天气系统能够找到更多与厄尔尼诺类似的模式，或者学会预测龙卷风或季风，远比人类做得更好。把海量的气象数据填到一个表单中来让人们理解是很困难的。与此相反，我们的智能天气系统却能直接感知和思考天气。

利用其他各种大型分布式传感系统，我们可以构建出很多智能系统，用于了解和预测动物迁徙、人口结构变化以及疾病传播，等等。假如我们有全国电网的传感器，与这些传感器连接的智能机器将能观察到电量使用的涨落，就像我们在公路上看到的交通涨落，以及在机场看到的人流涨落。通过反复接触，人们能够学会预测这些模式，这一点你只要问问路上来往车辆中的司机或者机场保安就可以知道。类似地，我们的智能电网监视器将能够预测电力需求，或者可能导致停电的危险情况，远比人做得要好。我们也许还能综合天气传感器和人口学传感器，用来预测政治动乱、饥荒或疾病暴发。就像一个超级聪明的外交家，智能机器可以在减少冲突和人类痛苦方面发挥重要的作用。你可能会认为智能机器需要情感才能预测人类的行为模式，我不这么认为。我们并不是天生就有文化、价值观和宗教信仰的，我们是通过学习得到这些的。正是因为我能够通过学习来理解有不同价值观的人的动机，所以即使智能机器本身没有这些情感，它也能够理解人类的动机和情感。

我们还可以通过感官来感知微小实体。理论上，我们可以通过传感器获取细胞或大分子中的模式。例如，当前一个重要的挑战是要了解如何从构成蛋白质的氨基酸序列来预测蛋白质分子的形状。一旦能够预测蛋白质的折叠和相互作用，将会加速许多疾病的药物和治疗的研制。工程师和科学家们已经建立了蛋白质的三维可视化模型，正努力预测这些复杂分子的行为。很多尝试已经证明这个任务太困难了。而配置了专门感官的超智能机器也许能够解答这个问题。如果这听起来很牵强，那么你可以想想，假如人类能够解决这个问题，我们并不会感到惊讶。我们无力解决这个问题的主要原因可能在于，人类感官与我们要了解的物理现象之间的不匹配。智能机器可以定制感官，而且记忆规模超过人类，因此它能够解决这个问题，而我们不能。

通过适当的感官和皮层记忆的细微调整，我们的智能机器可以在数学和物理的虚拟世界中生活和思考。例如，在数学和科学的很多探索中，需要理解超三维世界中对象的行为方式。研究空间本身性质的弦论，需要思考有着10个甚至更多维度的宇宙。人类很难思考四维或更高维的数学问题。也许，正确设计的智能机器可以用你我理解三维空间的方式去理解高维空间，从而可以预测这些空间中的行为方式。

最后，我们也许可以将一堆智能系统融合形成庞大的层次结构，正如我们的大脑皮层会联合听觉、触觉和视觉，形成更高级的皮层体系。这样的系统将自动学习建模并预测多个智能机器产生的思维模式。利用因特网等分布式通信媒介，我们可以将各个智能机器分布在全球各地。更庞大的层次体系可以学会更深层的模式，看到更复杂的类比。

以上这些思考的重点是想表明，我们有很多激动人心的方式可以让类人脑机器超越人脑。智能机器也许思考和学习都比我们快100万倍，能记住海量的详细信息，或者能看懂非常抽象的模式。智能机器会有比我们更灵敏的感官系统，或者会有分布式的感官系统，或者会有能够感知非常微小的现象的感官系统。它们可能会在三维、四维或更高维度的

空间中思考。所有这些有趣的可能性都不是智能机器对人类行为的刻板模仿，也不涉及复杂的机器人技术。

图灵测试将智能等同于人类行为，现在我们已经能充分感受到，图灵测试曾多么地限制我们对未来的憧憬。通过首先理解智能是什么，我们可以制造出智能机器，这远比仅仅复制人类的外在行为更有价值。我们的智能机器将是惊艳世人的工具，将极大扩展我们对宇宙的认识。

* * *

这一切需要多长时间才能成为现实呢？我们需要多久才能造出智能机器，50年、20年，还是5年？在高科技领域有个说法，你的短期期望往往需要更长的时间，而长期期望则往往比你的预期发生得快。这种情形我见证过很多次。有人曾经在会议上宣布一项新技术，并宣称它将在4年内进入千家万户。这个预测最后被证明是错的。4年变成了8年，当人们开始觉得它永远不会普及时，当整个想法看上去无路可走时，它突然开始一飞冲天并引起巨大轰动。同样的事情也可能会发生在智能机器上。最初的进展将会很慢，然后才迅速发展起来。

在神经科学的会议上，我喜欢在房间里四处走动，让每个人预测要过多久我们才能掌握大脑皮层的工作原理。不到5%的人会说“永远不会”或者“我们已经掌握了”（对于以神经科学研究为生的人而言这是个惊人的回答），还有5%的人说5—10年，有一半的人说10—50年或者“我的有生之年”，剩下的人说50—200年或者“我有生之年看不到了”。我站在乐观主义这一边。几十年来我们的研究一直处在“缓慢”的阶段，所以在很多人眼里，理论神经科学和智能机器的进展已经彻底停滞。根据过去30年取得的进展来判断，可以很自然地假设我们仍旧距离答案很远。但我相信我们正处在一个转折点上，这个领域即将起飞。

我们有可能加速到达未来，也有可能将转折点移到离现在更近。这本书的目标之一就是要说服你相信通过正确的理论框架，我们可以在理

解大脑皮层方面迅速取得进展。以记忆-预测框架作为指导，我们可以破译大脑的工作原理和人类思考方式的细节。这正是我们制造智能机器所需要的知识。如果这是正确的模型，我们就能够迅速地推进。

所以，虽然我很难预测智能机器时代何时成为现实，但是我认为，如果现在有足够多的人在尝试解决这个问题，我们也许在短短几年内就能造出有用的原型系统和大脑皮层模拟系统。我希望在10年内，智能机器能够成为最热门的科学技术领域之一。我无法说得更详细，因为我知道，孕育重大变革所需的时间往往容易被低估。那么，为什么我会这么看好理解大脑和制造智能机器的发展速度呢？我的信心主要来自于我在智能问题上长年花费的心血。1979年，当我第一次喜欢上大脑这个研究问题时，我觉得解决智能谜题有可能在我的有生之年实现。多年以来，我仔细考察了人工智能的衰落，神经网络的兴衰，以及19世纪90年代“脑的十年（Decade of the Brain）”工程。我已经看到了人们对于理论生物学，特别是理论神经科学的态度是如何演变的。我已经看到了关于预测、层次表征和时间的思想，是如何悄悄地进入神经科学的语言中的。我已经看到了我本人和同事们对这个问题的理解不断进步。18年前，我对预测的作用感到非常兴奋，多年来我一直在用很多方式测试它。由于我已经在神经科学和计算机领域浸淫了超过20年的时间，我的大脑可能已经建立起了关于科学技术变革的高层模型，这个模型预测了智能机器的快速发展，而现在就是转折点。

结语

天文学家卡尔·萨根（Carl Sagan）曾说，对一件事物的理解并不会减损它的奇妙与神秘。许多人担心对日常事物的科学理解会使其神奇感消失，仿佛知识会榨取生活的滋味与色彩，但我同意萨根的说法。事实上，正是因为有理解，我们才越发适应自身在宇宙中的角色，同时宇宙也显得愈发神秘缤纷。作为无限宇宙中一枚充满活力，拥有意识、智能

和创造力的微尘，远远要比生活在一个处于小宇宙中心的、单调有限的地球上有趣得多。同样，了解我们大脑的工作方式，并不会削弱宇宙、生活以及未来的神奇和神秘。当我们运用这些知识来认识自身、建造智能机器，继而获得更多的知识时，我们只会愈发惊叹不已。

因此，了解大脑和建造智能机器，是人类值得为之付出努力同时也是势在必行的追求。

我希望这本书能够吸引一批年轻的工程师和科学家们去研究大脑皮层，并利用记忆-预测框架来建造智能机器。人工智能在其鼎盛时期曾是一场声势浩大的运动，它曾拥有自己的期刊、学位课程、书籍、商业计划和企业。神经网络在20世纪80年代蓬勃发展时也同样激荡人心。然而人工智能和神经网络的学科框架并不适用于建造智能机器。

现在，我建议我们朝向一条更有前途的新道路前进。如果你正在读高中或大学，而本书激励了你去研究这项技术——打造第一款真正意义上的智能机器、助力于开启一个行业——那么我鼓励你放手去做，并努力让它成为现实。创业成功的关键一点是：你必须在百分之百地肯定自己能成功之前就投身于新的领域。时机很重要，如果起跳太早，你会备受煎熬，而如果等待不确定性完全消散，往往就为时已晚。我坚信，现在正是开始设计和建造仿大脑皮层记忆系统的最好时机。这一领域在科学和商业上都将变得极为重要。在未来10年，建立在层级结构记忆系统上的新产业内将会出现如“英特尔”和“微软”一样的巨头企业。这样规模的尝试，在经济上有一定的风险，对智力也有很高的要求，但无论如何是值得放手一搏的。我希望你能同我和其他愿意接受挑战的人们站在一起，开创有史以来最伟大的技术。

---

[1] Remember to fell Mary that the bog is ready to be piqued up. 这是为了模拟语音识别的错误，对应的语音可能是：Remember to tell Mary that the boy is ready to be picked up. 记得告诉玛丽这个小男孩已经准备好等人去接了。——译者注

附录 可检验的11个预测

每个理论都会引出可验证的假设，因为实验验证是确定新想法有效性的唯一可靠方法。幸运的是，记忆-预测框架以生物学为基础，由它所引出的一些具体的新预测是能够接受实验验证的。在本书的附录中，我列出了其中的一些预测，它们能够证实或证伪本书中所提出的观点。这部分所涉及的内容比第六章的内容还要艰深，并且对于理解此书的其他部分来说，并不是必须要掌握的。其中有几个预测，由于涉及对刺激出现的期望和假设，因此只能在清醒的动物或人类被试身上进行验证。以下所列预测在重要性上不分先后。

预测1

在包括初级感觉皮层在内的大脑皮层的各个区域中，我们都应该能找到这样一类细胞：它们因预测感觉事件的发生而表现出兴奋反应，但对感觉事件本身并不反应。

例如，冷泉港实验室的托尼·扎多（Tony Zador）发现，大鼠初级听觉皮层中有一些细胞会在它预期听到一个声音而实际上并没有声音时激活。这应该是大脑皮层的一个普遍特性。我们应该能在视觉皮层和躯体感觉皮层找到类似的预期反应。细胞对一种感觉输入的信息形成预期并激活，这就是假设的定义，亦是记忆-预测框架的基本前提。

预测2

一个预测在空间上越具体，那些因预期这一事件而兴奋的细胞就越贴近初级感觉皮层。

如果使用视觉模式序列来训练一只猴子，让它对一个出现在确定时间的特定视觉模式产生预期，当它开始期待预期模式时，我们就应该能观察到相应细胞的活动增强（对预测1的重申）。如果猴子学会了预测一张面孔，但不确定是什么样的面孔以及何时出现的话，我们就应该期望在人脸识别区域找到与预期有关的细胞，而不是在更低一级的视觉区域。但是，如果猴子注视目标，并学会预期在其视野中的精确位置上出现的特定模式的话，那么，我们就应该在V1或附近找到与预期有关的细胞。表征预测的神经活动沿着大脑皮层的层级结构一直向下传递，传递的远近取决于该预测的特性。有时它可以一路传至初级感觉区，有时它又会停在较高层级的区域。其他感觉通道应该也存在着类似的情况。

预测3

那些因期待感觉信息的输入而表现出兴奋活动的细胞，应该位于皮层的第2、第3和第6层，而预测应该在此层级结构的第2、第3层停止向下传递。

预测是这样在层级结构中传播的：先是经由第2、第3层细胞，然后投射到第6层。第6层中的细胞穿过下一皮层区域层级结构中的第1层细胞，再激活另一组第2及第3层细胞，等等。因此我们应该在这些层（第2、第3和第6）的细胞上观察到由预期而引发的兴奋活动。我们曾提过，在第2和第3层中的活跃细胞表征为一组潜在的兴奋细胞垂直柱，即潜在的预测。第6层中的活跃细胞由一组数量更少的细胞垂直柱所表征，代表着大脑皮层某一区域的具体预测。当预测在层级结构中向下传播时，兴奋活动将最终止步于第2和第3层。举例来说，假设一只老鼠学会了预期两种不同音调中的一种。根据外部线索，老鼠能够知道它何时

会听到这两个音调，然而不能预测会是哪一种。这时，我们应该会看到由期待而产生的兴奋活动出现在第2或第3层中同时表征这两个音调的细胞垂直柱上。因为老鼠并不能预测将要听到的是哪一个音调，所以在同一区域的第6层应该不会出现兴奋活动。如果在另一项实验中，老鼠能够准确预测它将要听到的音调，那么我们就应该会在第6层中对应该特定音调的细胞垂直柱上观察到兴奋活动。

我们不能完全排除在第4和第5层也存在预测细胞的可能。在这些层中很可能存在着好几类功能未知的细胞。因此，这个预测相对来说较弱，但我觉得还是值得一提的。

预测4

在第2和第3层中应该有这样一类细胞，它们优先接收来自更高层皮层区域的第6层细胞的输入信息。

记忆-预测模型的一部分内容提到，所习得的同时发生的模式序列将形成一个暂时的恒定不变的表征，即我称为“名称”的东西。我认为，这个“名称”就是大脑皮层某个区域的第2或第3层细胞中以不同垂直柱表示的一组细胞。只要有属于该序列的事件在发生，这些细胞就会保持兴奋状态（例如，只要某首旋律中的某个音符被听见，就会有一组细胞保持活跃）。代表这一序列名称的这组细胞，是由大脑皮层更高层区域中第6层细胞的反馈所激活的。我认为这些名称细胞是在第2层，因为它们邻近第1层。但它也可能是第2和第3层中任一类有树突连接到第1层的细胞。该命名系统要想正常运行的话，这些名称细胞的顶树突必须优先与位于第1层而源于皮层较高区域第6层细胞的轴突形成突触连接。它们应当避免同源于丘脑细胞的轴突形成突触。因此，我们的理论认为，在第2和第3层中应该存在着一类细胞，它们的顶树突位于第1层，并倾向于同源自更高皮层区域的第6层细胞的轴突形成突触。而在第1层形成突触

的其他细胞则不应有这种倾向。在我看来，这是一个很强的新预测。

由此预测所推出的一个必然结论是：在第2和第3层中应当存在着另一类细胞，其顶树突会优先与来自丘脑非特异性区域的轴突形成突触。这些细胞能够预测序列中将要出现的下一个项目。

预测5

预测4中所提到的“名称”细胞，在已习得序列出现时应当始终保持兴奋状态。

一组在已习得序列出现时保持兴奋的细胞，就是一个可预测序列的“名称”。因此，即便垂直柱中其他部分（在第4、第5、第6层）的细胞兴奋状态发生了变化，我们仍应能观察到一些细胞组仍然保持在兴奋状态。不幸的是，我们并不知道这些“名称”细胞的兴奋活动看起来是什么样子。举例来说，一个名称模式的恒定兴奋状态可以非常简单，可以是这组名称细胞集体一致地发出单个动作电位。而这样一组兴奋细胞可能很难被检测到。

预测6

在第2和第3层中还应存在着另一类细胞（与预测4和预测5中所提到的“名称”细胞不同），它们会对预料外的输入信息产生兴奋，而对于预料内的输入信息不作反应。

这一预测背后的依据是：预料以外的事件必须在皮层系统中向上传递，而当一个事件是预料之中时，我们不会再将它原封不动地往上传递，因为它在本处已经被预测到了。因此，在第2和第3层中应该要有一

类细胞，它们与预测4和5中所提到的“名称”细胞不同。这类细胞会对预料之外的事件产生兴奋反应，而对预料之内的事件则不作反应。这些细胞的轴突应该投射到皮层中的较高层区域。我提出了一种机制，来实现这类细胞两种兴奋状态间的变化。这类细胞可以通过由名称细胞激活的中间神经元来抑制，但目前对此机制还无法得出一个一致的预测。我们仅仅知道，一些细胞应该表现出不同的兴奋活动。在我看来，这又是一个很强的新预测。

预测7

预测6中提到，预料以外的事件应该在皮层体系中向上传递。事件越是罕见新颖，输入就会传递得越高。完全新颖的事件应该能传递到海马体。

学习得滚瓜烂熟的模式，在较低层的脑区就会得到预测，反之，输入模式越新颖，就应当在皮层体系中传递得越高。我们应该能够设计出一个实验来捕捉这种差异。比如，可以让某人听一首并不熟悉但简单的旋律。如果他听到某个音符，虽说这个音符有些出乎预料，但它与整个音乐的风格是一致的，那么这个预料之外的音符就会引起听觉皮层活动的变化，并且这种变化会向上传递至皮质体系中的某一较高层级。然而，如果他听到的并不是一个与音乐风格一致的音符，而是一个完全没有意义的声音，如东西破碎的声音，我们将会预期由此声音所引发的活动变化，会沿着皮层系统传递到更高的层级。如果被试期待听到的破碎声，而实际上听到的是一个音符，那么结果就应该相反。此预测可以通过功能性磁共振成像（fMRI）技术在人类被试的大脑中加以验证。

预测8

顿悟会引发一串精准的预测性兴奋，沿皮层体系向下传递。

当你终于弄清一个令人费解的感觉模式，比如对图12中的斑点狗的识别，你恍然大悟地发出“哦”的惊叹声时，大脑皮层的某个区域正尝试着用一个新的记忆匹配它所接收到的输入。如果在此区域完成了匹配，那么预测将会快速持续地向下传递至皮层体系中的所有下级区域。如果这是对刺激的正确辨识，那么体系中的每个区域将接连不断地迅速确定正确的预测。当观看拥有两种理解的图像时，也会产生同样的效果，比如一个可以被看成两副面孔的花瓶轮廓或是一个内克尔立方体（一个以不同朝向交替呈现的立方体的图像）。每当对此类图像的知觉发生变化时，我们都应当看到新的预测沿皮层体系向下传递。在最低层的区域，比如V1，表征图像中一条线段的垂直柱无论在图像的哪一感知过程中都应当一直保持兴奋（假设眼睛没有移动）。然而，我们可能会看到该垂直柱中的一些细胞处于交替兴奋状态。也就是说，虽然图像中的低级特征相同，但对图像的不同理解，会使得同一垂直柱中兴奋的细胞不同。最主要的一点是，当高层次的知觉发生变化时，我们应该能看到新的预测沿皮层体系向下传递。

对已知视觉对象的每一次扫视，都会引发类似的预测流传递。

预测9

在记忆-预测模型中，锥体神经元能够探测到薄树突上同时发生的突触输入。

多年以来，人们一直以为神经元可能是简单的集成单元，它们整合所接收的全部突触输入信息，来决定自身是否要发放。今天的神经科学对神经元工作方式的理解仍然存在着很多盲点。一些人仍然坚持视神经元为简单的集成单元。许多神经网络模型就是基于这样的认识而建立

的。还有许多关于神经元的模型假设，神经元在工作时它的树突的每个部分都是独立运作的。在记忆-预测模型中，神经元在一个很短的时间窗内只需要探测少许一致的突触兴奋。该模型甚至只需一个足以引起细胞发放的突触就能运作，但更可能的情况是，在薄树突上会有两个或更多邻近的兴奋突触。因此，一个拥有成千上万个突触的神经元就能够学习对许多独立不同的输入模式准确地做出反应。这并不是什么新的想法，已经得到了实验证据的支持，只是它从根本上背离了已使用多年的标准模型。如果神经元被证明并不对精确而独立的输入模式产生兴奋发放的话，那么记忆-预测理论的完整性将很难维持。只有薄树突上的众多突触需要以这种方式工作，在厚树突上或细胞体附近的突触并不需要。

预测10

对外界输入信息的表征，将伴随训练在皮层体系中逐渐下移。

我认为，通过反复训练，大脑皮层会在其较低层级区域重新学习模式序列。根据模式序列的记忆将会如何改变传递至更高级皮层区域的输入模式，自然可以推断出这一观点。这个过程会有几种结果，其中之一是：经过大量的训练后，我们应该在大脑皮层较低区域中找到对复杂刺激产生反应的细胞，而在经过最低限度的训练后，我们应该会在大脑皮层较高区域中找到相应的细胞。例如，在人的大脑中，经过对单个字母的识别训练，我会预期在IT区找到对印刷字母产生反应的细胞。然而，在完成对整个单词的识读之后，我会预期在V4区的不同部分而不是IT区找到对字母产生反应的细胞。使用其他物种、其他脑区和其他刺激进行此实验，应该也会得到相似的结果。这个学习过程中的另一个结果是，产生回忆和检测错误的地方应该发生变动。也就是说，经过高度学习的模式感觉沿皮层体系向上传播的距离应该比较短。这一点能够通过成像技术来检验。我们也应该能检测到对某些刺激反应时的变化，因为输入

信息不需要在皮层中传递太远就能得到识别并回忆起来。

预测11

恒定表征应该存在于所有的皮层区域。

大家已经知道，存在着这样的细胞，它们有选择性地只对不随细节变化的输入信息作出反应。科学家们已经发现了对面孔、手以及比尔·克林顿等等产生反应的细胞。记忆-预测模型预言，大脑皮层的所有区域都应形成恒定表征。这些恒定表征应能对某一皮层区域以下的所有感觉通道产生反应。例如，如果在我的视觉皮层中有一个“比尔·克林顿”细胞，那么每当我看到比尔·克林顿时，它就应该会兴奋。如果在我的听觉皮层有一个“比尔·克林顿”细胞，那么每当我听到“比尔·克林顿”这个名字的时候，它也应该兴奋起来。于是，在同时接收视觉和听觉输入信息的联合皮层区，也应该存在着一些细胞，无论是看到比尔·克林顿或是听到这个名字，都会让它们产生兴奋。恒定表征应该存在于所有的感觉通道，甚至是运动皮层中。在运动皮层区，细胞会表征复杂的运动序列。运动区内的级别越高，表征就越复杂也越恒定。（最近的研究发现，猴子大脑中存在着对从手到口的动作产生反应的细胞。）这些都不是什么新的预测。大多数研究人员都相信，大脑皮层的许多部分都有恒定表征形成。然而，尽管我将此视为事实，但这些恒定表征是否无处不在尚未得到证实。记忆-预测模型预言，我们在大脑皮层的每一区域都能找到这样的细胞。

* * *

以上所列的预测是可以验证本书中提到的记忆-预测模型的一些方式，当然肯定也会有其他方式。然而，我们只能证明某个理论是错的，而不可能证明某个理论是正确的。因此，即便以上所有预测都被证明是

正确的，也无法证明记忆-预测模型的假设一定正确。但这能作为该理论的强有力的支持证据。反而言之，如果上述预测中有一些被证明是错误的，也并不一定就让这整篇论述无效。对于某些预测来说，也存在其他替代方法可以实现所需的行为。例如，还有其他的方法能够用来创建序列的名称。本附录的目的只为表明，我们的模型可以引出几个预测，而这些预测又都可以被检验。设计实验是一项具有挑战性的工作，并不在本书所应涵盖的范围之中。利用成像技术如功能性磁共振成像来验证这一理论将会是不错的办法。目前已经有很多脑成像实验室，与单细胞记录技术和记录相比，在脑成像实验室验证这些假设相对要更快一些。

参考书目

大部分科学书籍和期刊文章后面都会附上冗长的参考书目，这既是为了列出文中所引的出处，也是为了协助读者更好地理解。由于本书旨在面向大众，包括那些以前没有神经科学知识的读者，我尽量避免了学术风格的写作。此参考书目同样旨在协助非专业读者了解更多与本书主题相关的信息。在书目中，我没有全部列出已发表的相关研究，也没有全部列出对此领域的重大发现有所贡献的个人。我列出的是一些自选书目，相信它们对于那些有兴趣了解更多有关大脑知识的读者会是很好的自学材料。书目中还列出了一些我认为对专家们有参考价值的文献。此外，你可以在互联网上找到许多相关主题的深入讨论。更多书目资源可在这本书相关的网站www.OnIntelligence.org上获取。

很遗憾，在这些书目里，你不会看到太多介绍大脑整体理论的资料，就像我在序言里所说的那样，对这一主题尚没有多少教科书般的资料。而对于本书中所列出的具体论点，参考资料就更少了。

人工智能和神经网络的历史

［1］Baumgartner, Peter, and Sabine Payr, ed.*Speaking Minds: Interviews with Twenty Eminent Cognitive Scientists.*Princeton, N. J.: Princeton University Pr, 1995

这本书包含了对人工智能、神经网络和认知科学领域内众多重要思想家的有趣访谈，可以作为读者了解智能研究的历史及其精神的一份轻松有趣的概要。

［2］Dreyfus, Hubert L.*What Computers Still Can't Do: A Critique of Artificial Reason.* Cambridge, Mass.: MIT Pr, 1992

这本书包含对人工智能的严厉批判。最早一版的书名叫作*What Computers Can’t Do*，数年后以修订后的书名再版。它是由最有力的批评家写就的一部人工智能的深入历史。

［3］Anderson, James A., and Edward Rosenfeld, ed.*Neurocomputing, Foundations of Research.*Cambridge, Mass.: MIT Pr, 1988

这部大书是1890至1987年间神经网络和大脑理论的重要论文集，书中的论文按时间顺序排列并带有注解。其中包含了W.S. 麦卡洛克（McCulloch），W. 皮茨（Pitts），唐纳德·赫布（Donald Hebb），史蒂夫·格罗斯伯格（Steve Grossberg）和其他许多人的论文，每篇论文前都附有编辑的介绍。此书能够为那些想要了解这一领域重要历史文献的读者提供一条便捷的途径。

［4］Searle, J. R. “Minds, Brains, and Programs.”*The Behavioral and Brain Sciences.* 1980（3）：417～24

这本书中提出了著名的“中文屋”论证，反对将计算作为智能模型。你也可以在互联网上找到许多关于塞尔的这个思维实验的介绍和讨论。

［5］Turing, A. M. “Computing Machinery and Intelligence.”*Mind.*1950（59）：pp. 433～60

书中提出了用来检测机器是否拥有智能的著名的“图灵测试”。有关图灵测试的大量参考资料和讨论，同样可以在互联网上找到。

［6］Palm, G ünther.*Neural Assemblies: An Alternative Approach to Artificial Intelligence.*New York: Springer Verlag, 1982

对自－联想记忆的了解有助于理解皮层的工作方式和它如何存储序列模式。虽然介绍自－联想记忆的书籍已有不少，但我还没有发现任何一本书能够简单易懂地总结那些我视为重要的知识。帕尔姆是这一领域的先驱，他的这本书很难找，而且读起来也不容易，但它涵盖了自－联想记忆包括序列记忆的基础知识。

大脑皮层和普通神经科学

以下书籍推荐给那些想要了解更多有关神经生物学和大脑皮层知识的读者。

［1］Crick, Francis H. C. “Thinking about the Brain.”*Scientific American*：1979（241）：pp. 181～88. Also available*in The Brain A Scientific American Book.* San Francisco: W. H. Freeman, 1979

这就是那篇让我开始对大脑产生兴趣的论文。虽然时隔25年之久，这篇由弗朗西斯·克里克写就的论文在我看来依旧鼓舞人心。

［2］Koch, Christof.*Quest for Consciousness: A Neurobiological Approach.* Denver, Colo.: Roberts and Co., 2004

每年都有那么几本大众性的大脑介绍书籍出版，克里斯托夫·科赫的这本书探讨的是意识，但它实际上涵盖了大部分与大脑、神经解剖学、神经生理学以及意识相关的内容。如果你想从一本可读性较高的书开始了解有关神经生物学和脑科学的基本知识，它将会是一个很好的选择。

［3］Mountcastle, Vernon B.*Perceptual Neuroscience: The Cerebral Cortex.* Cambridge, Mass.: Harvard University Pr, 1998

这是一本致力于介绍有关新大脑皮层的一切知识的好书。文字优美，布局清晰，虽然有些技术派的风格，仍不失为是阅读上的享受，是学习大脑皮层最好的入门书籍之一。

［4］Kandel, Eric R., James H. Schwartz, Thomas M. Jessell, ed.*Principles of Neural Science,*4th ed. New York: McGraw-Hill, 2000

这是一本涵盖了所有与神经有关的知识的大型百科全书。不太适合睡前阅读，但作为一本参考书是极好的。它为神经系统的所有部分都提供了详细的介绍，包括神经元、感觉器官和神经递质。

［5］Shepherd, Gordon M., ed.*The Synaptic Organization of the Brain, 5th ed.*New York:Oxford University Pr, 2004

这本书对我颇有助益，虽然我更喜欢早期那个只有一位作者的版本。它对我理解大脑的所有部分，尤其是突触，提供了技术上的帮助。我用它当参考书。

［6］Koch, Christof, and Joel L. Davis, ed.*Large-scale Neuronal Theories of the Brain.*Cambridge, Mass.: MIT Pr, 1994

有关大脑整体理论的书非常之少。这一本正是关于这一主题的论文汇编，虽然书中的大部分的论文并没有达到书名所提出的目标。这本书概述了人们为理解大脑的整体工作方式所采取的各种方法。你会在书中多处发现记忆－预测框架的零星片段。

［7］Braitenberg, Valentino, and AlmutSchu?z.*Cortex: Statistics and Geometry of NeuronalConnectivity,*2nd ed. New York: Springer Verlag, 1998

这本书描述了小鼠大脑的统计特性。我知道这听起来并不怎么令人兴奋，但它的确是本让人耳目一新的有用的书。它用数据描述了大脑皮层。

神经科学的专业论文

以下论文是本书所提及的一些重要概念的原始出处，其中绝大部分都只能在大学图书馆或网上找到。

［1］Mountcastle, Vernon B. “An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System.”in Gerald M. Edelman and Vernon B. Mountcastle, ed.,*The Mindful Brain.*Cambridge, Mass.: f MIT Pr, 1978

我在这篇论文中第一次阅读到蒙卡斯尔关于整个新大脑皮层如何基于一个共同原则工作的论述。蒙卡斯尔也曾提出将皮质的垂直柱作为基本计算单位的想法。这些想法为本书所提出的理论提供了前提和灵感。

［2］Creutzfeldt, Otto D. “Generality of the Functional Structure of the Neocortex.”*Naturwissenschaften*. 1977（64）：pp. 507～17

我在写完《论智能》后才接触到这篇论文。它同蒙卡斯尔的论文一样，认为存在着一个通用的皮层算法。此篇论文的发表时间稍早于蒙卡斯尔的论文，对后者也是一个很好的补充。

［3］Creutzfeldt, Otto D. “Generality of the Functional Structure of the Neocortex.”*Naturwissenschaften*. 1977（64）：pp. 507～17

这是一篇描述视觉皮层层级结构的现代经典论文。而记忆-预测模型建立在这样一个假设上：不光是视觉系统，整个大脑皮层都具有层级结构。

［4］Sherman, S.M., R.W. Guillery. “The Role of the Thalamus in the Flow of Information to the Cortex.”*Philosophical Transactions of the Royal Society of London*. v. 357, no. 1428（2002）：1695～708

这篇论文提供了关于丘脑组织的概述，并提出了谢尔曼－吉勒里假说，该假说认为丘脑负责掌控皮层区域之间的信息流。我在第六章“沿体系向上传递的另一途径”部分详细阐述了这一想法。

［5］Rao, R. P., D. H. Ballard. “Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-field Effects.” *Nature Neuroscience*. v. 2, no. 1（1999）： 79～87

我将此篇论文加入书目，作为最近讨论预测和层级的一个例子。饶和巴拉德提出了一个皮层层级中的反馈模型，其中较高区域中的神经元会去预测较低区域的活动模式。

［6］Guillery, R. W. “Branching Thalamic Afferents Link Action and Perception.”*Journal of Neurophysiology*. 2003（90）：539～48

［7］Young, M. P. “The Organization of Neural Systems in the Primate Cerebral Cortex.” *Proceedings of the Royal Society: Biological Sciences*. 1993（252）：pp. 13～18

这两篇写作精良的论文所提供的证据表明，运动行为和感官知觉密切相关，并且是同一进程的两个部分。吉勒里认为，所有的感觉皮层都在运动行为中发挥作用；杨的论文表明，运动皮质和躯体感觉皮层联系得非常紧密，它们应该被视为同一个系统。我在第六章中简要探讨了这些想法。

致谢

每次有人问起“你做什么工作？”时，我总是不知该如何作答，因为我实际上并没有做什么具体的工作。然而，我身边的人们似乎做出了许多出色的成果，我所扮演的角色不过是时不时地激发一下他们的斗志，并在必要时为整个团队指出新的道路。如果一定要说我在职业生涯中有过成功的话，那么它应该主要归功于团队伙伴们的辛劳和智慧。

我曾有幸拜访过许多科学家，他们几乎每一个都令我受益匪浅。对于此书中的理论，他们亦有贡献。虽然下面只提到了少数几位，但我希望向他们所有人表示最深的谢意。

在红木神经科学研究所（RNI）和美国加州大学戴维斯分校就职的布鲁诺·奥尔斯豪森（Bruno Olshausen）就像是一个活的神经科学百科全书，他不断地为我指出知识上的欠缺，并为补救我的无知提出了许多好的建议，这对我来说是无价的帮助。同样在RNI就职的比尔·索弗基（Bill Softky）是第一个教给我皮层层级结构中的时间缩减以及薄树突特性等知识的人。加州大学尔湾分校的里克·格兰杰（Rick Granger）帮助我深入了解了序列记忆以及丘脑的潜在作用。美国加州大学伯克利分校的鲍勃·奈特（Bob Knight）和加州理工学院的克里斯托夫·科赫（Christof Koch）在红木神经科学研究所的建立和许多其他科学问题上给予了我极大的帮助。RNI的所有员工激励并督促着我完善书中的这些想法。这本书中的许多观点都直接来源于在RNI所进行的会议和讨论。衷心感谢大家。

唐娜·杜宾斯基（Donna Dubinsky）和埃德·科利（Ed Colligan）是我长达十几年的商业合作伙伴。在他们的辛勤工作和无私帮助下，我实

现了既当企业家又兼职大脑理论研究这样一种非同寻常的状态。唐娜常说，她的目标之一就是使我们的业务获得成功，这样才能让我花费更多的时间在大脑理论上。如果不是唐娜和埃德，这本书就不会存在。

离开众人的帮助，我无论如何也写不出《智能时代》。我的经纪人吉姆·莱文（Jim Levine）在我还不确定要写什么的时候，就已经对它信心十足——如果你没有像吉姆这样棒的经纪人，千万不要写书。吉姆把我介绍给了本书的合著者，桑德拉·布莱克斯利（Sandra Blakeslee）。我希望这本书的受众多、可读性高，而桑迪的参与对实现这一目标至关重要。如果书中有任何艰难晦涩的部分，责任一定在我。桑迪的儿子，科普作家马修·布莱克斯利（Matthew Blakeslee）为我提供了书中所使用的几个例子，“记忆-预测框架”这一名称也是他提议的。与亨利·霍尔特（Henry Holt）的诸位共事是非常愉快的经历。在此特别感谢亨利·霍尔特的总裁和出版人约翰·施特林（John Sterling）先生。我与约翰只见过一次面，在电话里交谈过几次，而就是这几次短短的交流，对本书的结构产生了巨大的影响——他很快就想到了我提出一个新的智能理论将会面临的问题，然后向我建议这本书应该如何写、如何定位。

我要感谢我的女儿，安妮和凯特，当爸爸在电脑键盘旁度过许多个周末时，她们没有大哭大闹。最后，我要感谢我的妻子——珍妮特。做我的妻子并不轻松，我爱她多过于爱大脑。